Australian Bureau of Statistics

6401.0.60.003

Information Paper

Making Greater Use of Transactions Data to Compile the Consumer Price Index, Australia

Australia

2016

Information Paper

Making Greater Use of Transactions Data to Compile the Consumer Price Index, Australia

Australia

2016

David W. Kalisch
Australian Statistician

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30AM (CANBERRA TIME) TUES 29 NOV 2016

ABS Catalogue No. 6401.0.60.003

Produced by the Australian Bureau of Statistics

INQUIRIES

- For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070.

CONTENTS

PREFACE

The environment in which statistical agencies operate is changing. New opportunities to access and interrogate big data sets are becoming available, increasing the potential to provide new insights into matters of importance. The statistical landscape is becoming more complex, expectations of decision makers are growing, and National Statistical Offices (NSOs) are being challenged to deliver the best possible statistical program in more efficient and innovative ways. To remain relevant and to capitalise on these opportunities of the dynamic information environment, the Australian Bureau of Statistics (ABS) has undertaken a research program to enhance the Australian Consumer Price Index (CPI). This publication focuses on one component of that program, that being to maximise the use of transactions data to compile the Australian CPI (ABS 2015).

While transactions data provide opportunities to enhance the CPI, it also creates methodological challenges for price index practitioners. To date only a handful of countries have implemented transactions data for use in their CPI, each using different methods and practices. This is in part due to a lack of consensus amongst leading NSOs on the best aggregation method when utilising large volumes of data, as well as differences in circumstances of each NSO in producing their CPI.

This publication presents the case for change to some of the methods used to compile the Australian CPI. The methodological changes outlined in this publication will enhance the accuracy of the CPI by utilising big data sources. This publication examines the emerging academic literature relating to big data methods and supplements this literature review with ABS research. This publication concludes by outlining a path towards implementation. Of note, some technical challenges remain unresolved. User and stakeholder input is sought to contribute to the resolution of these challenges. A final ABS publication will be released on this topic in mid-2017 which will outline the precise methods to be implemented in the CPI.

The ABS would like to acknowledge the input and advice provided by Professor Jan de Haan of Statistics Netherlands and Delft University of Technology during the preparation of this publication.

David W. Kalisch
Australian Statistician

ABBREVIATIONS

ABS	Australian Bureau of Statistics
ABS DQF	Australian Bureau of Statistics Data Quality Framework
CPD	Country Product Dummy
CPI	Consumer Price Index
EA	elementary aggregate
EC	Expenditure Class
FHES	Furnishings, Household Equipment and Services
GEKS	Gini, Eltetö and Köves, and Szulc
GK	Geary-Khamis
HES	Household Expenditure Survey
ILO	International Labour Organization
ITRYGEKS	Imputation Törnqvist Rolling Year GEKS
MEI	main economic indicator
NSO	National Statistical Office
QAUV	Quality Adjusted Unit Value
SKU	Stock Keeping Unit
TPD	Time Product Dummy
UNECE	United Nations Economic Commission for Europe

INTRODUCTION

BACKGROUND

1.1 The Australian Bureau of Statistics (ABS) has a long history, over 100 years, of providing trusted official statistics to inform decisions on issues of importance to Australia. However, the environment in which the ABS operates is changing. New opportunities to access and interrogate big data are becoming available, increasing the potential to provide new insights into matters of importance to Australians. The statistical landscape is becoming more complex, expectations of decision makers are growing, and the ABS is being challenged to deliver the best possible statistical program in more efficient and innovative ways.

1.2 The launch of barcode scanner technology in Australia during the 1970s, and its growth in the 20th century, has enabled retailers to capture detailed information on transactions at the point of sale. Transactions data is high in volume and contains detailed information about individual transactions or summaries, date, quantities, product descriptions, and values of products sold. As such it is a rich data source to NSOs that can be used to enhance their statistics, reduce provider burden, and reduce associated costs of physically collecting data.

1.3 From March quarter 2014 the ABS significantly increased its use of transactions data to compile the Australian CPI (ABS 2013), now accounting for approximately 25% of the weight of the Australian CPI. The approach adopted was a 'direct replacement' of observed point-in-time prices with a unit value calculated from the transactions data. This unit value is then combined with other price observations at the elementary aggregate (EA) level. The EA forms the lowest level of CPI aggregation, and is used to calculate price indexes for expenditure classes (EC) and the All groups CPI. The major benefit of this approach compared to the traditional point-in-time pricing is that a more representative price for each product is used to compile the CPI.

1.4 While this has enhanced the Australian CPI, it is acknowledged that more can be done with transactions data to compile official statistics than traditional approaches. The availability of timely expenditure information (for weighting purposes) enables the calculation of weighted bilateral indexes (such as the Fisher and Törnqvist) that account for consumer substitution across time. However, traditional methods have known fundamental weaknesses when chaining price indexes at a high frequency (Ivancic, Fox and Diewert 2011).

1.5 New methods and processes are required to maximise the use of transactions data. Typically, multilateral index methods have been used in the spatial context to compare price levels across different regions, however academics and NSOs are proposing they be used to make price comparisons across multiple (three or more) time periods. Temporal multilateral methods produce weighted price indexes and have the property of transitivity[1].

1.6 In recent years there has been an increase in the range of multilateral methods proposed for use in CPI aggregation when using transactions data. Earlier research conducted by the ABS assessed well-known multilateral methods in a temporal context, including the Gini, Eltetö, Köves and Szulc (GEKS) and the Time Product Dummy (TPD) which were proposed by Ivancic, Fox and Diewert (2009). Since then however, a number of new methods have been proposed for producing multilateral temporal indexes, such

1 Where the price change measured between two time periods is independent of whether they are compared directly or via some other period. Chained weighted price indexes do not share this property.

BACKGROUND *continued*

as the multilateral method recently implemented by Statistics Netherlands for mobile phones adapted from the Geary-Khamis (GK) method (Chessa 2016).

1.7 Practical challenges exist when applying these methods in the CPI. When a multilateral method is extended by an additional period (e.g. quarter), previous price movements are revised, which is unacceptable for NSOs. Additionally, the length of time the multilateral method uses price and expenditure information for index compilation is important in order to account for seasonal availability of product prices without impairing the index. There are a number of approaches available to NSOs to deal with these challenges, and these will be discussed and supported empirically in this publication.

1.8 While multilateral methods have strong theoretical and practical properties, only a handful of NSOs have actually implemented them[2], displaying caution in implementation and a divergence in methods and practices. The choice of index formula is contextual: the best elementary aggregation method depends, in part, on the transactions datasets that are available to the NSO. Other considerations specific to the NSO, such as the environment in which they operate, and the methods they use to compile the CPI more broadly are also necessary. In order to consider multilateral methods in an Australian context, this publication will use the ABS Data Quality Framework (DQF) together with empirical evidence to demonstrate support for a multilateral aggregation method that is suitable for the Australian CPI.

CURRENT ABS APPROACH USING TRANSACTIONS DATA

1.9 The current ABS approach to compile the CPI using transactions data is consistent with the International Labour Organization (ILO 2004), and is a replacement of directly observed (point-in-time) prices with a unit value calculated from the transactions data. The unit value approach takes expenditure and quantity data by product over the period of interest (e.g. quarter) to calculate an average unit price. It allows for better outlet coverage as unit values are calculated over all of a respondent's outlets, rather than just a sample. The major benefit of this approach compared to the traditional point-in-time pricing is that unit values provide a more accurate summary of an average transaction price than an isolated price quotation (Diewert 1995).

1.10 The chosen time aggregation is important when using transactions data, with more time aggregation providing more stable estimates of price change as price and quantity bouncing behaviour is smoothed out. Unit prices are derived at both monthly and quarterly frequencies from transactions data to ensure elementary aggregation is consistent across other modes of price collection. The ABS aggregates both transactions data and field collected prices at the EA level using the geometric mean formula.

1.11 When using the unit value approach, all of the difference between successive unit values for the same item is attributed to price (i.e. it assumes quality differences are zero). Items must therefore be tightly defined at a fine level of aggregation to maximise homogeneity and prevent quality differences from affecting the unit values. This can be challenging when using transactions data as these datasets tend to exhibit a high level of churn. The ABS defines products as homogenous by using product classifications provided by Australian proprietors known as the stock keeping unit (SKU).

2 See Appendix 1 for a list of NSOs using transactions data in the compilation of the CPI.

CURRENT ABS APPROACH USING TRANSACTIONS DATA *continued*

1.12 The ABS has developed an approach to make explicit quality adjustments to products that have changed using the detailed descriptions provided in the transactions data. There are broadly three main scenarios which initiate the quality adjustment process for prices obtained from transactions data.

i) A new SKU is brought into the sample as a replacement for another SKU to which it is not directly comparable: this requires calculating a previous period price for the new item.

ii) A SKU in the sample has a quantity change (e.g. packet size): this requires calculating a quality adjustment factor (i.e. ratio of packet sizes) by extracting information for the item description. This quality adjustment factor is applied to product expenditures prior to calculating a unit price.

iii) A SKU is replaced with a similar SKU, possibly with a quantity change: This also requires calculating a quality adjustment factor by extracting information for the item descriptions. To capture price change due to products of the same quality with different SKUs (e.g. product relaunches), the ABS has developed a method to link comparable new and disappearing products. The linking process uses information on the product description, price, expenditures, timing (when products appear or disappear on sales listings) and quantity sold.

1.13 To ensure the sample derived from transactions data remains relevant, the ABS has incorporated an approach to review the sample each period to identify items that have become less relevant based on expenditure shares, and recommend suitable replacements. Existing SKUs sampled in the CPI are assessed in each period against several relevancy tests to identify items that have shown large drops in expenditure. The main principle behind these relevance tests is that the items have a stable expenditure share within the CPI product group (i.e. items which fail these tests are recommended for replacement), with an output file recommending replacements for analysis. The current approach to determine product replacements is manually driven, with the strategy being intentional and manageable with the current sample sizes.

1.14 The relevancy tests described above are different for seasonal items (exclusively seasonal fruits) where certain varieties are popular in one part of the year, but not available at all in other periods. For these seasonal items, the ABS selects the highest expenditure variety of the product each period. This approach ensures the most representative product is selected for that period, which to avoid selecting clearance or unsuitable products, is subject to a minimum monthly expenditure threshold. This is a continuation of the field collection practice, where field officers collect the price of the most representative variety each period.

1.15 Reference prices are calculated for any new items that are introduced to the sample via the replacement process. To ensure the new products are introduced into the sample at a normal selling price, the ABS calculates a 'normal' unit value price (which has an average level of specialling) as the reference unit value price. This is done by first calculating an average discount over the previous 12 months and then applying this average discount to the highest price in the previous month/last month of previous quarter. This is to reflect a 'normal level of discounting' and ensure the index is not distorted by unusual price activity in the previous period.

CURRENT ABS APPROACH USING TRANSACTIONS DATA *continued*

1.16 Summarising the current approach, the calculation of average unit prices over both time and respondent outlets provides a more representative price paid by consumers over the reference period (e.g. month, quarter) than a point-in-time price from a specific retailer. The replacement strategy also informs sampling decisions, where high expenditure items dictate the products sampled in the CPI. While the implementation of the direct replacement method using unit values has provided the Australian CPI with a significant enhancement, it is recognised that more can be done with transactions data.

BENEFITS AND CHALLENGES IN USING MULTILATERAL METHODS

1.17 Big data is becoming available in new ways, providing NSOs with opportunities to deliver statistical output in a more efficient and innovative manner. Traditional price aggregation methods based on weighted price index formula are suitable for use when the basket of items remains fixed over time. Moving to transactions data, items are dynamic, appearing and disappearing on a regular basis over time. In order to accommodate a dynamic universe of items these techniques try to match as many of these as possible and form a continuous price index where price movements are chained together. Frequent chaining of weighted bilateral price indexes causes a chain drift[3] problem due to the nature of price and quantity bouncing. This breakdown of bilateral price index formula has led researchers and NSOs to investigate alternate methods for price index aggregation.

1.18 The consideration of multilateral methods is motivated by a number of studies showing that bilateral methods can suffer from drift when chained at high frequencies (Ivancic, Fox and Diewert 2011; van der Grient and de Haan 2011). While a chained bilateral index measures short term movements accurately, a comparison between published index levels several periods apart may not accurately reflect the price change over that time. Multilateral methods are designed to preserve transitivity; they are independent of the choice of base period. However, as new data becomes available over time, multilateral methods revise previously estimated index numbers which is troublesome in the context of producing a non-revised CPI. There are several strategies available to NSOs to extend a multilateral index which are discussed later in this publication.

1.19 Since multilateral methods use data from several periods, an unavoidable consequence is that the price comparison between any two periods depends on prices and weights in other periods as well, which could affect the relevance of the index. A tension between 'characteristicity' - the relevance of a price comparison to the periods under consideration - and 'chain drift' - loss of transitivity - must be considered when assessing multilateral methods (Ivancic, Fox and Diewert 2011). Empirical analysis should assist with this decision, with particular interest focusing on differences in the direction of the index movement, or the timing of turning points, that result from using different extension methods.

1.20 By making greater use of the number of price observations from the dynamic transactions dataset, sample representativeness in the product dimension would be increased, the probability of sampling error would be reduced and biases (e.g. item substitution, and new and disappearing items) associated with purposive sampling would be eliminated, hence the accuracy of the index would be enhanced. This is in

3 Chain drift is the failure of an index to return to parity after prices and quantities revert back to their original values.

BENEFITS AND CHALLENGES IN USING MULTILATERAL METHODS *continued*

contrast to the traditional approach which is based on the purposive sampling technique, where a sample of prices are taken across several dimensions, including products, respondents, locations and time (Mackin, Oehm and Gow 2012). The incorporation of transactions data in to the Australian CPI has increased the representativeness of the sample in both the location and time dimensions. However, making greater use of transactions data would further enhance the CPI. In addition, using expenditure data to weight the dynamic universe of prices based on their economic importance would make the price index movement a more accurate reflection of the concept that the CPI measures: the price change facing Australian households for a basket of goods and services (ABS 2011).

1.21 Increasing the number of price observations from transactions data and aggregating by way of a multilateral method would reduce the number of resources required at various stages of the statistical cycle (i.e. sampling, collection and processing of prices). The opportunity to automate processes to reduce costs associated with producing a CPI is attractive to all NSOs. An added advantage for the Australian CPI is that efficient methods also facilitate more timely outputs - processes that require less manual intervention can be completed sooner after the reference period, and make the production of higher-frequency outputs more feasible.

1.22 Utilising transactions data could also require an increase in resources at new points of the statistical cycle since new data usually requires new processes to be followed. The mapping of items to CPI classifications is currently a labour intensive process at the ABS as each respondent has unique classification structures. This process of mapping is required every time new transactions data is secured, or when there is a high amount of product churn in the datasets. The ABS plans to further investigate automated mapping in the future to facilitate increased use of transactions data in a less resource intensive manner.

1.23 While multilateral methods have several attractive properties, they are more complex than bilateral methods. It is important that NSOs are able to explain their properties to stakeholders so they can assess their suitability for their purposes. Additionally, it is important that NSOs explain published movements using these methods. Tools that can decompose price movements at the lower levels of aggregation are required to enable this explanation to occur.

BACKGROUND

2.1 Multilateral methods possess a number of desirable qualities, both theoretical and practical, to produce temporal price indexes from big data sources. The theoretical qualities include maintaining transitivity when reweighting and chaining frequently; while, from a practical perspective, automated processes allow a greater sample of products to be used to produce price indexes.

2.2 The international price statistics community has reached a consensus that multilateral methods are the most appropriate approach to produce temporal price indexes when using big data (UNECE 2016). At this point in time no specific multilateral method has received international endorsement. Over recent years researchers have been proposing several multilateral methods for producing temporal indexes. As described by de Haan (2015, p.1), "in spite of all the research that has been done, so far only a handful of countries have actually implemented scanner data into their CPI [...] using different methods and practices." This reticence in implementation and divergence in methods and practice stems from a number of factors, including a lack of consensus amongst researchers or leading NSOs about the best method to use, and the circumstances in which each NSO produces their CPI.

2.3 The difference between the multilateral methods themselves is in the aggregation approach each takes. For instance, the TPD method uses a regression based approach that estimates price change over time by measuring the statistical relationship between prices, products and time. The GK and Quality Adjusted Unit Value (QAUV) methods in general use a unit value approach to estimate price change and so necessitate product homogeneity to be able to standardise the quantities into common units. The difference between the QAUV_TPD and GK is in the approach to standardising the quantities – these will be discussed in this section. The GEKS method applies a geometric mean of the ratios of all bilateral indexes.

2.4 This section describes the multilateral methods the ABS is investigating for future implementation in the CPI. The choice of index formula, in part, depends on the transactions datasets that are available to the NSO. The transactions data used in this publication[4] contain the following generic data fields: city, date, SKU, product description, proprietor classifications, quantity and value sold of each product. Due to the lack of detailed product characteristics available in the data, this publication focuses on a selection of popular matched-model multilateral methods including the TPD, GK, QAUV and GEKS methods[5].

TPD

2.5 The TPD method uses a regression approach that is similar to hedonic based methods previously used in the Australian CPI - it uses the statistical relationship between prices, products and time to directly estimate price change over time. The TPD method is a technique adapted by Aizcorbe, Corrado and Doms (2003) from the Country Product Dummy (CPD) method for spatial comparisons.

2.6 The TPD model is estimated by pooling together data for a specified window length (T+1) and modelling the log of price against time and product binary indicators. The TPD model is expressed as:

4 See Attachment 1 of ABS (2015) for a full list of ECs where transactions data is available.
5 See de Haan, Willenborg and Chessa (2016) for a more exhaustive list of multilateral methods available for temporal aggregation.

TPD *continued*

$$\ln p_i^t = \alpha^0 + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t \quad (2.1)$$

where,

$\ln p_i^t$ = log of price for item i in period t

α^0 = intercept term

δ^t = time parameter corresponding to time period t

D_i^t = time dummy variable, equal to 1 if the price observation p_i^t pertains to period t and 0 otherwise

γ_i = product parameter corresponding to product i

D_i = product dummy variable, equal to 1 if the price observation p_i^t pertains to item i and 0 otherwise

ε_i^t = error term

2.7 The time effect δ^t reflects the overall price level in period t relative to a reference period 0, while the product effect γ_i reflects the typical price of product i relative to the reference product N. Transforming equation 2.1, the predicted prices are $\hat{p}_i^0 = \exp(\hat{\alpha}^0)\exp(\hat{\gamma}_i)$ and $\hat{p}_i^t = \exp(\hat{\alpha}^0)\exp(\hat{\gamma}_i)\exp(\hat{\delta}^t)$ for all products belonging to an EC. Using the ratio of these predicted prices, the price index can be directly estimated from the modelled time effect parameters as follows:

$$P_{TPD}^{0,t} = \frac{\hat{p}_i^t}{\hat{p}_i^0} = \exp(\hat{\delta}^t) \quad (2.2)$$

where,

$P_{TPD}^{0,t}$ = price movement between periods 0 and t

$\hat{p}_i^0$ = predicted price of product i from period 0

$\hat{p}_i^t$ = predicted price of product i from period t

2.8 The model specified in equation 2.1 does not produce a weighted price index using ordinary least squares estimation. In order to produce a weighted price index, expenditure shares for each product i and time period t are derived and used in fitting a weighted least squares version of equation 2.1, such that the following sum is minimised:

$$S = \sum_{U} s_i^t (e_i^t)^2 \quad (2.3)$$

where,

S = weighted sum of squared residuals

U = set of all price observations in the window

s_i^t = expenditure share of product i relative to other products sold in time period t

e_i^t = residual error term of price observation p_i^t

QAUV AND GK

2.9 The QAUV and GK methods both appeal to the notions of homogeneity and unit values by expressing the quantity of products into common units, then calculating a unit value across all products. The QAUV and GK have different methods of standardising quantities, these are discussed in more detail below.

2.10 The QAUV method can be described as constructing an implicit price index using information on value and quantity available from transactions datasets. The idea behind the QAUV method is to separate products into relatively homogeneous groupings, and calculate implicit price indexes by dividing a value index by a quantity index. This is expressed as:

$$P^{0,t} = \frac{V^{0,t}}{Q^{0,t}} \quad (2.4)$$

where,

$P^{0,t}$ = unit value index between periods 0 and t

$V^{0,t}$ = value index between periods 0 and t

$Q^{0,t}$ = quantity index between periods 0 and t

2.11 While $Q^{0,t}$ in equation 2.4 is constructed across a group of homogeneous products, it is likely that some differences in quality between individual products remain. In order to produce a standardised quantity index, each item within the homogeneous grouping can be standardised with respect to an arbitrary base item, then aggregated to produce a quantity aggregate. As a result, the standardised quantity index can be expressed as:

$$Q_{QAUV}^{0,t} = \frac{\sum_{i \in U^t} v_{i/b} q_i^t}{\sum_{i \in U^0} v_{i/b} q_i^0} \quad (2.5)$$

where,

$Q_{QAUV}^{0,t}$ = standardised quantity index

$v_{i/b}$ = adjustment factor comparing item i to base item b

q_i^t = quantity of item i in period t

q_i^0 = quantity of item i in period 0

U^t = sample of products from period t

U^0 = sample of products from period 0

Substituting the quantity index derived in equation 2.5 into equation 2.4, the QAUV method can be expressed as:

$$P_{QAUV}^{0,t} = \frac{V^{0,t}}{Q_{QAUV}^{0,t}} = \frac{\sum_{i \in U^t} p_i^t q_i^t / \sum_{i \in U^t} v_{i/b} q_i^t}{\sum_{i \in U^0} p_i^0 q_i^0 / \sum_{i \in U^0} v_{i/b} q_i^0} \quad (2.6)$$

QAUV AND GK *continued*

2.12 The expression in equation 2.6 does not explicitly describe a suitable method to calculate the adjustment factors $v_{i/b}$ .This publication investigates two methods for calculating adjustment factors. Firstly, the method proposed by Chessa (2016) is investigated which is a temporal adaption of the GK method used to construct spatial price indexes, where adjustment factors are estimated using an iterative approach based on two sets of simultaneous equations. The adjustment factor is the quantity weighted average of deflated prices and can be expressed as:

$$v_{i/b} = \sum_{z \in T} \varphi_{i,z} \frac{p_{i,z}}{P_{GK}^{z}} \quad (2.7)$$

where,

$v_{i/b}$ = weighted average of deflated prices of product i

$\varphi_{i,z}$ = quantity share of product i in time period $z \in T$

$p_{i,z}$ = price of product i in time period $z \in T$

P_{GK}^{z} = price index in time period $z \in T$

The quantity share component from equation 2.7 is derived as:

$$\varphi_{i,z} = \frac{q_{i,z}}{\sum_{s \in T} q_{i,s}} \quad (2.8)$$

where,

$q_{i,z}$ = quantity of product i in period z

$\sum_{s \in T} q_{i,s}$ = total quantity of product i over all time periods $s \in T$

In this publication, we denote the use of quality adjustment factors as defined above for the QAUV method as GK.

2.13 An alternative option for calculating adjustment factors is the TPD approach outlined in de Haan (2015). The TPD method uses the regression model specified earlier in equation 2.1, which estimates the predicted prices $\hat{p}_i^0 = \exp(\hat{\alpha}^0)\exp(\hat{\gamma}_i)$ and $\hat{p}_i^t = \exp(\hat{\alpha}^0)\exp(\hat{\gamma}_i)\exp(\hat{\delta}^t)$ for all products. In order to express quantities in constant units, the predicted prices for all products can be compared to a base item, such that the adjustment factor becomes the following:

$$v_{i/b} = \frac{\hat{p}_i^t}{\hat{p}_b^t} = \frac{\exp(\hat{\gamma}_i)}{\exp(\hat{\gamma}_b)} \quad (2.9)$$

where,

$\hat{\gamma}_i$ = product parameter corresponding to product i

$\hat{\gamma}_b$ = product parameter corresponding to product b

The TPD approach is denoted in a way to form adjustment factors for QAUV as QAUV_TPD.

GEKS

2.14 The GEKS method takes the geometric mean of the ratios of all bilateral indexes (calculated using the same index number formula) between a number of entities. For spatial indexes these entities are generally countries, while for price comparisons across time, the entities are time periods.

2.15 The bilateral index formula chosen for this publication is the Törnqvist index which can be expressed as:

$$P_T^{0,t} = \prod_{i=0}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{\frac{s_i^t+s_i^0}{2}} \quad (2.10)$$

where,

$P_T^{0,t}$ = Törnqvist index between periods 0 and t

p_i^t = price of item i in period t

p_i^0 = price of item i in period 0

$\frac{s_i^t+s_i^0}{2}$ = average expenditure share item i across periods 0 and t

n = number of matched items between periods 0 and t

2.16 The GEKS is calculated as the geometric mean of the ratios of all matched-model bilateral indexes $P^{l,t}$ and $P^{l,0}$ where each period l is taken in turn as the base (de Haan 2015). The GEKS method can be expressed as:

$$P_{GEKS}^{0,t} = \prod_{l=0}^{T} \left(\frac{P^{l,t}}{P^{l,0}}\right)^{\frac{1}{T+1}} = \prod_{l=0}^{T} \left(P^{0,l} \times P^{l,t}\right)^{\frac{1}{T+1}} \quad (2.11)$$

where,

$P_{GEKS}^{0,t}$ = GEKS index between periods 0 and t

$P^{l,t}$ = Törnqvist index between periods l and t

$P^{l,0}$ = Törnqvist index between periods l and 0

In this publication we denote the use of the Törnqvist bilateral formula to form GEKS indexes. This is referred to as GEKS indexes in the following sections.

BACKGROUND

3.1 When a multilateral method is used to produce a temporal index, each bilateral price comparison depends on prices observed in other periods of the multilateral comparison window. As a result, incorporating a new period into the multilateral comparison window may alter the price comparisons of earlier periods. Ivancic, Fox and Diewert (2009) observe that this creates a revision problem, since published price indexes are only revised in exceptional circumstances.

3.2 This section describes a selection of methods available to extend a multilateral temporal index when a new period of data is available to form a continuous non-revised CPI.

3.3 The choice of length for the estimation and splicing windows has generally defaulted to 13 months (for monthly series) or 5 quarters (for quarterly series), in order to account for any seasonality in product prices (Ivancic, Fox and Diewert 2011). A recent paper by de Haan (2015) indicates 25 months or 9 quarters to be a more appropriate window length. Empirical results from ABS research suggest that two years and one period - 25 months or 9 quarters - should be used for the estimation window length. This is the basis for the decision to use this as the window length in this publication. The exception to using a 25 month or 9 quarter rolling window is the direct extension method in effort to test methods described in Chessa (2016).

DIRECT EXTENSION

3.4 The direct extension method constructs a price series in a similar manner to the approach used to produce direct bilateral indexes. When data is available for a new period, the direct extension method recalculates the multilateral comparisons and publishes the index level corresponding to the new period. The direct extension method is currently being used by Statistics Netherlands in CPI compilation for mobile phone products (Chessa 2016), with the direct multilateral index chained annually (in the December month), so only the price observations since that last December contribute to each direct index calculation. This application of the direct extension method for a given year can be expressed as:

$$P_D^t = P^l \times P^{l,t} \quad (3.1)$$

where,

P_D^t = index level in period t

P^l = index level in the latest chaining period l before t (e.g. the previous December)

$P^{l,t}$ = price movement between periods l and t, using a multilateral window starting in period l and ending in period t

MOVEMENT SPLICE

3.5 The movement splice method also involves incorporating a new time period into a new multilateral comparison window and extending the index based on a price comparison from this new window. The index level in this new period is calculated by multiplying the previous (published) index level by the price movement between the previous and the new period, as estimated using the new multilateral window (Ivancic, Fox and Diewert 2011). This is analogous to the approach used to produce chained bilateral indexes and can be expressed as:

$$P_{MS}^t = P^{t-1} \times P^{t-1,t}(t-T) \quad (3.2)$$

MOVEMENT SPLICE *continued*

where,

P^t_{MS} = index level in period t

P^{t-1} = index level in the previous period

$P^{t-1,t}(t-T)$ = price movement between $t-1$ and t using the latest multilateral window between $t-T$ and t

WINDOW SPLICE

3.6 The window splice method, proposed by Krsinich (2016), uses the same rolling window approach to extend the index when a new period of data is available. However, the movement and window splice methods use price movements from the latest multilateral comparison window to update the index in different ways. Whereas the movement splice method joins the last period-on-period movement from this window, the window splice method joins on the latest full window onto the index level of T periods earlier. This can be expressed as:

$$P^t_{WS} = P^{t-1} \times \frac{P^{t-T,t}(t-T)}{P^{t-T,t-1}(t-T-1)} \quad (3.3)$$

where,

P^t_{WS} = index level in period t

P^{t-1} = index level in the previous period

$P^{t-T,t}(t-T)$ = price movement between $t-T$ and t using the latest multilateral window between $t-T$ and t

$P^{t-T,t-1}(t-T-1)$ = price movement between $t-T$ and $t-1$ using the previous multilateral window between $t-T-1$ and $t-1$

HALF SPLICE

3.7 The half splice method is a modification of the window splice, which involves splicing on an intermediate movement from the latest window - more than a single period-on-period movement, but less than the full movement. This splicing method has been suggested by de Haan (2015) and internally investigated by the ABS. The half splice method can be expressed as:

$$P^t_{HS} = P^{t-1} \times \frac{P^{t-\frac{T}{2},t}(t-T)}{P^{t-\frac{T}{2},t-1}(t-T-1)} \quad (3.4)$$

where,

P^t_{HS} = index level in period t

P^{t-1} = index level in the previous period

$P^{t-\frac{T}{2},t}(t-T)$ = price movement between periods $t-\frac{T}{2}$ and t using the latest window from $t-T$ to t

$P^{t-\frac{T}{2},t-1}(t-T-1)$ = price movement between periods $t-\frac{T}{2}$ and $t-1$ using the window from $t-T-1$ to $t-1$

HALF SPLICE *continued*

3.8 In this publication, we have denoted the combined use of a multilateral method in conjunction with an extension method by MM_EM, where:

- MM is the acronym for a given multilateral method (for example TPD, GK, QAUV_TPD, GEKS)
- EM is the acronym for a given extension method:
  - D - Direct extension
  - HS - Half splice
  - WS - Window splice
  - MS - Movement splice.

CRITERIA FOR ASSESSING MULTILATERAL METHODS

INTRODUCTION

4.1 This publication presents four multilateral methods for producing temporal indexes from transactions data, as well as four different methods for extending the index without revising it. It also notes that there is no broad consensus at present about the best choice of methods in this context. In the absence of such consensus, the ABS has developed a framework to guide the choice of method to use to make greater use of transactions data.

4.2 A natural starting point is the ABS Data Quality Framework (DQF) (ABS 2009), which prompts the consideration of statistical quality more broadly than focussing on a single aspect (e.g. accuracy). The framework includes seven dimensions of statistical quality:

- *Institutional Environment* - pertains to the institutional and organisational context in which a statistical producer operates
- *Relevance* - pertains to how well a statistic meets user needs
- *Timeliness* - pertains to how quickly and frequently the statistic is published
- *Accuracy* - pertains to how well a statistic measures the desired concept
- *Coherence* - pertains to how consistent the statistic is with sources of related information
- *Interpretability* - pertains to the information available to provide insight into the statistic
- *Accessibility* - pertains to ease of access to the statistic

4.3 The benefits and issues associated with using multilateral methods, as discussed earlier, can be linked to dimensions of this quality framework. These links are summarised in Table 4.1, and discussed further in the next few paragraphs.

TABLE 4.1: FRAMEWORK FOR ASSESSING MULTILATERAL METHODS

Consideration	*Quality dimensions*
Resources: does this method help facilitate more effective use of human and information resources?	Institutional Environment, Timeliness
Theoretical properties: what conceptual properties does the index method have, and how well do these align with the CPI purpose?	Accuracy
Transitivity: to what extent is the index transitive?	Accuracy, Coherence
Characteristicity: to what extent are price comparisons relevant to the time periods being compared?	Accuracy, Relevance
Flexibility: what scope is there to use or adapt the method for new statistical products or data sources?	Coherence, Institutional Environment
Interpretability: how easy is it to understand the method and the price movements it calculates?	Interpretability

4.4 This publication considers multilateral methods with the aim of maximising the use of transactions data in the CPI. As mentioned in the Introduction of this publication, these methods also offer opportunities for automating routine manual processes - making better use of human resources - and consequently making it more feasible to produce higher-frequency outputs. These considerations are linked to the *Institutional Environment* and *Timeliness* dimensions; they will be important as the ABS moves towards implementing a method for routine use in producing the CPI.

INTRODUCTION *continued*

4.5 The main advantage of maximising the use of transactions data is to improve the accuracy of the index, by reducing sampling error and biases associated with traditional sampling and weighting. However, the methods considered in this publication have different properties, both in theory and in practice. The next section explores the differences between the methods' theoretical properties and considers their advantages and disadvantages in a CPI context.

4.6 When proposing multilateral indexes for temporal aggregation, Ivancic, Fox and Diewert (2011) note a tension between transitivity - as chained bilateral indexes may drift over time - and characteristicity - as prices or preferences from distant periods may unduly influence multilateral comparisons. This links to the *Accuracy* dimension, but can also be seen as a trade-off between *Coherence* and *Relevance*, which depends more on the choices of window length and extension method than the multilateral method.

4.7 Other considerations relevant to multilateral methods include flexibility - how well a method can be adapted for new statistical products or data sources - and interpretability - how easy it is to understand a method in general as well as the index movements it produces. Both of these considerations are addressed in this section. *Interpretability* is already one of the quality dimensions, whereas flexibility links to *Coherence*, as well as *Institutional Environment* (as it reduces the potential complexity of systems).

THEORETICAL PROPERTIES

4.8 ILO (2004) assesses bilateral price indexes both from axiomatic/test (Chapter 16) and economic approaches (Chapters 17-18). The methods that emerge best from these assessments are the Fisher and Törnqvist indexes, which closely approximate one another in normal circumstances (ILO 2004: Chapter 17).

4.9 Similar approaches to assessing multilateral indexes in a spatial context have been developed and presented in several papers, especially Diewert (1999) and Balk (2001). These approaches reveal differences between methods; they are examined in detail for temporal applications in the next two sub-sections.

Test approach

4.10 Diewert (1999) and Balk (2001) propose similar sets of tests for spatial multilateral indexes. These Tests are expressed in terms of volume shares, which are equivalent to (normalised) multilateral quantity indexes. Table 4.2 contains the Tests proposed by both authors, expressed in terms of price and quantity comparisons between time periods. The adaptation of these Tests for the temporal context is discussed further in the Appendix 2.

4.11 Table 4.2 indicates which Tests are satisfied by each multilateral method considered in this publication, and whether the Test performance is preserved after the multilateral index is extended using the methods described earlier. The results for the GEKS and GK methods, as well as the TPD method for Tests 1 to 9, are taken from Balk (2001) and Diewert (1999); the other results are derived or discussed in the Appendix 2.

TABLE 4.2: TESTS FOR MULTILATERAL COMPARISONS

Test	*GEKS*	*TPD*	*GK*	*QAUV_TPD*	*Preserved after extension*
1 Positivity and continuity test: price and volume indexes are normalised, positive and continuous functions of (positive) prices and (nonnegative) quantities	Y	Y	Y	Y	Y
2 Weak proportionality test: if prices and quantities in all periods are proportional, price and volume comparisons depend only on those proportions (Balk only)	Y	Y	Y	Y	Y
2x If quantities in all periods are proportional, volume comparisons depend only on those proportions	Y	N	Y	Y	Y
2p If prices in all periods are proportional, price comparisons depend only on those proportions	Y	Y	Y	Y	Y
3 Homogeneity in quantities test: rescaling the quantities in some period does not alter the price comparisons if relative prices are unchanged	Y*	Y	N	Y	Y
4 Monetary units test: rescaling the prices in some period does not alter the volume comparisons if relative quantities are unchanged	Y	Y	Y	Y	Y
5 Commensurability test: changing the units in which all prices and quantities are measured does not alter the system of comparisons	Y	Y	Y	Y	Y
6 Symmetric treatment of entities test: reordering the periods does not alter the system of comparisons	Y	Y	Y	Y	N
7 Symmetric treatment of commodities test: reordering the commodities does not alter the system of comparisons	Y	Y	Y	Y	Y
8 Partitioning test: if there is a group of two or more periods with proportional prices and quantities, those proportions determine price and volume comparisons within the group, and aggregating price and quantities across periods within the group does not alter comparisons between periods outside the group	N	N	Y	N	?
9 Irrelevance of tiny periods test: as the aggregate volume in a period approaches zero, its influence on comparisons between other periods vanishes	N*	N	Y	N	Y
10 Monotonicity in quantities test: each period's volume share is an increasing function of its quantities	Y	?	N	?	?
11 Bilateral consistency in aggregation test: if we can group all periods into two groups such that prices and quantities in all periods in a group are proportional to a group-specific pair of reference price and quantity vectors, aggregate price and volume comparisons between groups are equal to Fisher price and quantity comparisons between the pairs of reference vectors (Diewert only)	Y	?	N	?	?
12 Additivity test: the system of comparisons is additive (Diewert only)	N	N	Y	Y	N

Note: * Balk (2001) considers a weighted GEKS, which satisfies test 9 but not test 3, whereas the opposite is true for the unweighted GEKS considered in this publication.

Test approach *continued*

4.12 If we first consider the performance of the multilateral methods without any extension. Table 4.2 reveals that no method satisfies all of these Tests. This leads both Diewert (1999) and Balk (2001) to conclude that the importance of different Tests, and hence the most appropriate method, depends on the situation.

4.13 Several Tests are satisfied by all four multilateral methods (1, 2, 4, 5, 6 and 7). Test 6 implies that the indexes are free from chain drift, which is the main motivation for using multilateral instead of chained bilateral indexes.

4.14 The Tests that differentiate the multilateral methods are of varying importance and relevance in the context of comparing prices across time. Arguably, the most interesting of these Tests are 3, 9 and 12.

4.15 The GK method's failure of Test 3 means that it gives greater weight to the price structures in periods with larger aggregate volumes. While Chessa (2016) argues that this is an advantage - because it gives out-of-season prices in seasonal product classes lower weight - it is not ideal to treat time periods unequally in temporal comparisons. Ensuring product classes are sufficiently broad that a substantial quantity of products are sold year

Test approach continued

round should both reduce the severity of the GK method's failure of this Test and reduce the risk that out-of-season prices have large weights. This also reduces the risk that there are any time periods with negligible volumes, and hence the other methods' failure of Test 9 should not be of great concern.

4.16 Test 12 is presented by de Haan (2015) and Chessa (2016) as an advantage of additive multilateral methods: they extend the notion of a unit value price to classes of broadly comparable products. If the reference prices used in the GK and QAUV_TPD methods adequately account for quality differences between products (that is, if they make products homogeneous) then this unit value approach seems appropriate. On the other hand, additive methods make strong assumptions about substitutability which are contentious from an economic perspective, as described in the next sub-section.

4.17 The monotonicity Test 10 is also interesting. The performance of several methods against this Test is unknown. A test for monotonicity in prices (which is not implied by monotonicity in quantities) may be more relevant for temporal price comparisons. Also, it is common for a product to be observed in only one period in a multilateral window, and it is not clear how monotonicity should apply in this scenario. This is an area for further study.

4.18 Of the remaining Tests, the TPD method's failure of Test 2x is not critical in this context, given it satisfies the corresponding price comparison Test 2p and the weak proportionality Test 2. Tests 8 and 11 both relate to how the system of multilateral comparisons behave when aggregated across or groups of similar entities are split up (e.g. countries in a bloc). NSOs are unlikely to aggregate or disaggregate time periods in this way. Conceivably, however, NSOs may wish to produce indexes at different frequencies (e.g. monthly, quarterly and annually), and it seems important that corresponding price movements, and trends over time, are broadly similar. An intermediate value test[6] may be more appropriate here.

4.19 Finally, note that extending these multilateral methods using the methods described earlier undermines their Test performance. This is because they involve splicing together price movements from successive multilateral windows, each of which makes use of slightly different price information. Of particular importance is that Tests 6 and 12 are prone to failure: i.e. extended indexes are not guaranteed to be free from chain drift, and extended additive indexes may only be approximately additive. Moreover, the differences between extension methods relate to how they deal with temporal ordering, both in constructing multilateral windows, and in combining certain price comparisons from successive windows to update the index. As such, different extension methods may yield different indexes: Krsinich (2016) and Chessa (2016) find substantial differences for some (but not all) datasets and product types. Empirical comparisons using data available to the ABS are presented in the next section.

Economic approach

4.20 The economic approach to justifying consumer price indexes assumes that consumers optimise their basket of purchases to maximise utility for a given budget, or minimise cost for a given level of utility. In this context, price (quantity) comparisons should reflect differences between optimal baskets of fixed utility (cost) in different time

6 For instance, that the (quarterly) price comparison between two quarters lies between the minimum and maximum price comparisons between any month in the first quarter and any month in the second quarter.

Economic approach continued

periods. This approach is addressed in more detail in the ILO CPI manual (2004: chapters 17 and 18).

4.21 An assumption underpinning the economic approach is that some function exists to express the utility of a basket of goods and services. In particular, this function describes the impact on utility of substitution between different products. Intuitively, if the price of one product grows more slowly than others, it may be advantageous for a consumer to direct a larger share of their budget towards that product over time.

4.22 Multilateral comparison methods assume that the utility function is uniform across all economic entities that are being compared. In the temporal context, there is a risk that this assumption fails due to changes in preferences over time. This risk can be mitigated by limiting the number of connecting time periods that are included in the multilateral window, though of course using too short a window increases the risk that Test 6 fails. This does not help to discriminate between methods.

4.23 However, the multilateral methods involve different assumptions about the shape or "functional form" of the utility function, and consequently the substitutability of different products. Diewert (1999) shows that the GEKS method is exact for a "flexible" functional form - that is, it expresses the price differences experienced by optimising consumers without imposing restrictive assumptions about how they can substitute between products. In contrast, the GK and other additive methods are consistent only with linear utility functions, so they may suffer substitution bias if consumer preferences are more complex. To the best of our knowledge, the TPD and QAUV_TPD methods have not yet been assessed rigorously from this economic approach.

4.24 Examining differences between the GEKS and other methods in the temporal context is an ongoing area of research. In practice, empirical testing does not reveal compelling evidence of substantial substitution bias in temporal GK indexes or other reference price indexes. Chessa (2016) finds little differences between methods. The ABS has used the method proposed by Hill (2000) to test for substitution bias in temporal GK, TPD and QAUV_TPD indexes but has found no clear evidence of such bias.

FLEXIBILITY

4.25 This publication deals primarily with methods for producing temporal price indexes from transactions datasets of a certain type that are available to the ABS. These datasets include information on quantities, expenditures, and descriptive information in a format that allows the identification and classification of products but makes it difficult to extract detailed product characteristics.

4.26 In evaluating index methods, it is an advantage for a method to be flexible enough to be used for a range of purposes and dataset types. If we prefer the properties of one method for aggregating transactions data, it makes sense to use variants of the same method for different data types and statistical products; this promotes coherence, aids interpretation, and reduces the potential complexity of processing systems.

4.27 In future, multilateral methods may facilitate the production of statistics the ABS does not currently produce, such as spatial indexes. This is feasible for any of the multilateral indexes considered in this publication, which are all based on spatial comparison methods.

FLEXIBILITY *continued*

4.28 The ABS may also acquire price datasets of other types, such as high frequency price datasets without weighting information (i.e. web scraped data) or transactions datasets that contain detailed product characteristics.

4.29 All of the methods described in this publication can be adapted for datasets without weighting information. This effectively treats each product as being of equal economic importance, which may be contentious; however, the question in this case is not whether the methods are capable of handling unweighted data, but whether it is appropriate to use unweighted data in the first place.

4.30 Finally, if transactions datasets contain detailed product characteristics, the GEKS, TPD and QAUV_TPD methods can be adapted to produce hedonic indexes using Time Dummy Hedonic indexes to substitute for the bilateral link formula (in the GEKS) or the TPD model (in the other methods). These hedonic indexes have the potential for better quality adjustment than the basic multilateral methods. In contrast, there is no established way of adapting the GK method to produce hedonic indexes, though Chessa (2016) suggests treating each combination of characteristics as a single product identifier. It is unclear whether this approach performs as well as a hedonic index in adjusting for quality change: this is an area of ongoing research.

INTERPRETABILITY

4.31 The ABS places a high value on transparency by understanding and explaining the statistics published, and describing and justifing the methods used. This is of critical importance for the CPI as a Main Economic Indicator (MEI). Two aspects of interpretability need consideration: first, to what extent the methods themselves are easy for index practitioners and users to understand; second, whether it is easy to understand the price movements each index produces, especially which products have the greatest influence on these movements and why.

Interpreting the methods

4.32 All of the multilateral methods considered here are more complicated than standard bilateral indexes. Nevertheless, all of them can be expressed in reasonably simple ways:

- The GEKS is perhaps the easiest to grasp - as it is based on traditional price index theory - as the multilateral movements are derived by combining superlative bilateral indexes.
- The TPD has a simple model representation, reflecting the relationship it models between prices, products and time; and as it is equivalent to the Rao method (Balk 2001; Rao 2005) it can also be presented using simultaneous equations or matrices.
- The GK and QAUV_TPD methods appeal to the notions of homogeneity and unit values: first, these methods calculate a set of adjustment factors to express all products in the same terms (make them homogeneous); second, they calculate a unit value index over all these products. The GK can be presented using simultaneous equations or matrices (see Collier 1999; Balk 2001). The QAUV_TPD is perhaps the most complicated because it involves calculating one set of price comparisons in estimating a TPD model, and then another set of price comparisons through the QAUV formula[7]. It is necessary to consider the trade-off between this complexity and the additivity that is induced by applying the QAUV formula.

7 De Haan and Krsinich (2014) and de Haan (2015) argue that the TPD and QAUV_TPD are approximately equal.

Interpreting the methods continued

4.33 Note that the extension methods may also affect interpretability: the movement splice and direct extension methods are perhaps easier to understand than the window or half-window splice methods. Ultimately, however, interpretability is somewhat subjective, and different methods may appear intuitive to different audiences.

Interpreting the index movements

4.34 Price movements can be analysed using a range of analytical techniques. A useful technique is to decompose the aggregate index movement into a simple function (sum or product) of contributions from individual products. This allows the identification of the products with the greatest influence on the index movement, which is useful for validating and explaining (and managing the quality of) the index movements.

4.35 Standard bilateral index movements are relatively easy to decompose, as they are expressed in terms of product prices and weights in two periods. In contrast, multilateral price movements depend on prices and weights across the multilateral window so they are more complicated to decompose.

4.36 In recent years, some progress towards the decomposition of multilateral indexes has been made. Van der Grient (2010) and de Haan and Hendriks (2013) show how both GEKS and TPD index movements can be decomposed into a bilateral Törnqvist term as well as a number of other factors. The ABS has further developed and tested methods to express GEKS and TPD movements as a product of contributions from individual items. These methods make it easy to identify the items and product groups with the greatest influence on index movements.

4.37 The existence of these decomposition methods is advantageous for the GEKS and TPD. However, as the QAUV and GK are additive methods, the implicit quantity index has a natural additive decomposition, and there may well be decompositions for the associated price indexes. This is an area for further research.

SUMMARY OF COMPARISONS BETWEEN METHODS

4.38 This section presents a framework for assessing multilateral methods, and explores differences between methods based on their theoretical properties, flexibility and interpretability.

4.39 In particular, the Theoretical Properties sub-section distinguished between methods on the basis of five factors:

- Whether they weight time periods equally or based on their volumes;
- Whether they are additive;
- Whether the price and quantity indexes are monotonic;
- Whether they produce consistent indexes at different frequencies;
- What assumptions they involve about uniformity over time and substitutability.

4.40 Operational decisions may reduce the differences between methods: for instance, constructing product classes of approximately consistent volumes means that all methods weight time periods approximately equally. Some properties, such as additivity, are not strictly preserved as the index is extended, and consequently their satisfaction should not be given too much weight in comparing methods. Other properties, such as monotonicity, require further study.

SUMMARY OF COMPARISONS BETWEEN METHODS *continued*

4.41 The Flexibility sub-section describes how all four methods can be used for new statistical products and data sources, noting that the GK method cannot be adapted to produce hedonic indexes but can make use of characteristic information in defining product identifiers.

4.42 The Interpretability sub-section argues that the four multilateral methods can be explained in reasonably simple ways, although they are all more complicated than bilateral index methods. It also notes the usefulness of movement decomposition methods for understanding index movements. Decomposition methods have been developed for GEKS and TPD methods but would require further work to develop for the GK and QAUV_TPD methods.

4.43 The ABS will consider these factors when assessing which is the most appropriate method for our purposes.

BACKGROUND

5.1 This section provides empirical results for the multilateral and extension methods described earlier. The intent behind the empirical analysis is to test the methods outside the theoretical realm, recognise characteristics that the methods treat differently, and compare how the methods behave at different compilation frequencies. A comparison of empirical results is important in the process of evaluating the available methods against the ABS DQF.

5.2 Empirical results are presented at both disaggregate and aggregate levels. The disaggregate level refers to price indexes below the published level that depend only on multilateral and extension methods for aggregation. These are presented to isolate the effects of different multilateral and extension methods. The disaggregate indexes are also compiled on a monthly frequency to observe how sensitive results are to a higher compilation frequency. The aggregate level refers to price indexes at the published level, which combine the disaggregate price movements of both transactions and non-transactions data respondents and aggregate using the Lowe index formula. The aggregate indexes are compiled on a quarterly frequency in an effort to produce results that are broadly comparable to the published CPI.

AGGREGATING PRICE MOVEMENTS

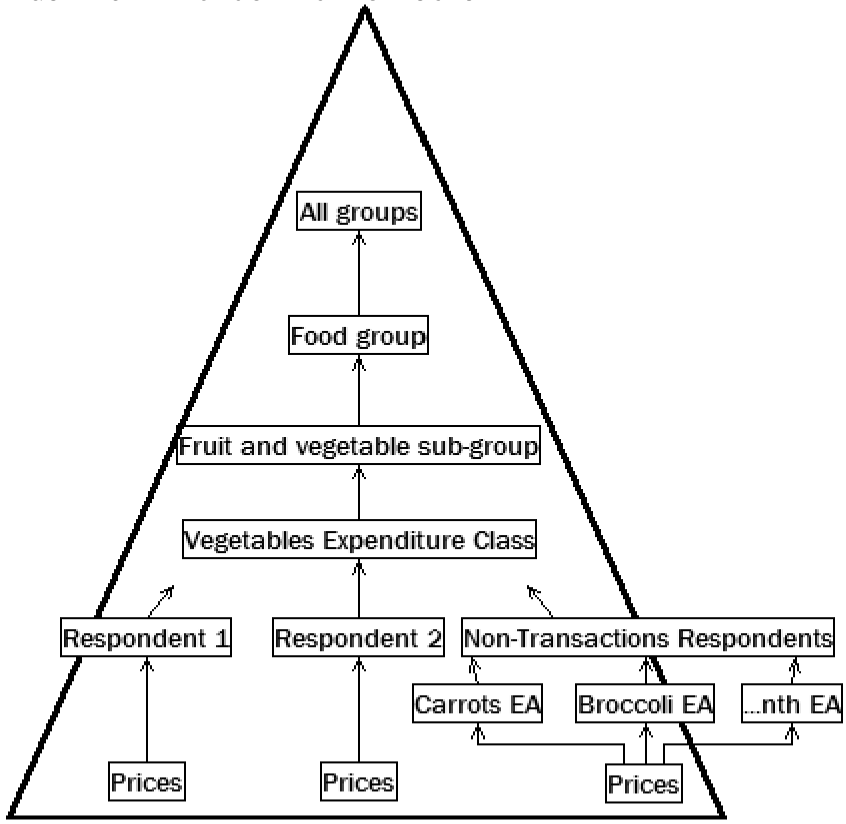
5.3 Traditional weighting practices vary at different levels of the CPI. At the EC level (and above), expenditure weights are currently updated every six years, primarily using Household Expenditure Survey (HES) data. Below the EC level, the expenditure weights are periodically updated to reflect households changing preferences. The compilation of elementary aggregates from individual prices uses an unweighted price index formula that considers each price observation of equal importance. The application of an unweighted formula is a practical approach due to data limitations where expenditure information at the item level is not available on a timely basis.

5.4 Transactions data provides expenditure information at the disaggregate level which opens up a range of index aggregation methods that more readily account for changes in household expenditure patterns. At the most disaggregate level, transactions data provide the ability to weight individual prices according to their level of expenditure which enhances the accuracy of outputs.

5.5 The approach adopted in this publication is to use the expenditure information available in transaction datasets across both individual items and respondents. The structure proposed in this publication is provided in Figure 5.1 for a subset of products in the All groups CPI. The proposed structure includes an additional level below the EC level that weights each respondent explicitly (by expenditure). Respondents can be weighted by using transactions data as well as other available data sources that contain information on each respondent's market share across various commodities.

AGGREGATING PRICE MOVEMENTS *continued*

FIGURE 5.1 PROPOSED CPI STRUCTURE



5.6 The proposed structure for transactions data respondents does not contain traditional elementary aggregates, instead mapping items by respondent directly to the EC level. This approach was chosen for two reasons. Firstly, transactions data is delivered to the ABS with various respondent specific classifications which make it difficult to map to the (approximately) 1000 elementary aggregates currently in the CPI. As a compromise, mapping respondent classifications to the EC level is a suitable alternative. Secondly, with a comprehensive mapping concordance constructed at the EC level, the accuracy of the price index will be enhanced since it is accounting for substitution across a wider range of items within each respondent. Moving forward, the ABS will further investigate the suitability of aggregating transactions data directly to the EC level.

5.7 With respect to non-transactions data respondents, at the EA level and above an explicit weight will be allocated in the index structure. The EA classifications and aggregation formula for non-transactions data respondents will remain the same as is currently used in the CPI.

EMPIRICAL RESULTS AT THE DISAGGREGATE LEVEL

5.8 Empirical results at the disaggregate level are presented below for a selection of ECs. Due to the sensitivities of reporting results below the published level, all results are expressed relative to an arbitrary multilateral/extension method.

5.9 With respect to reporting the behaviour of different multilateral methods, price indexes were firstly constructed on the entire dataset for a particular respondent spanning 58 time periods (period 0 to period 57). Across the majority of ECs, the multilateral methods yielded similar results. A selection of six ECs are presented below (Figure 5.2-5.7), which standardise each multilateral method (i.e. period 0 corresponds

EMPIRICAL RESULTS AT THE DISAGGREGATE LEVEL *continued*

to an index level of 100), and then express each multilateral method relative to the GEKS (e.g. TPD = TPD index less GEKS index). These figures below show the TPD, QAUV_TPD and GK typically oscillate within a two index point range of the GEKS, and the direction of this difference is typically mixed across the ECs.

FIGURE 5.2: SNACKS AND CONFECTIONERY EC (FULL WINDOW)

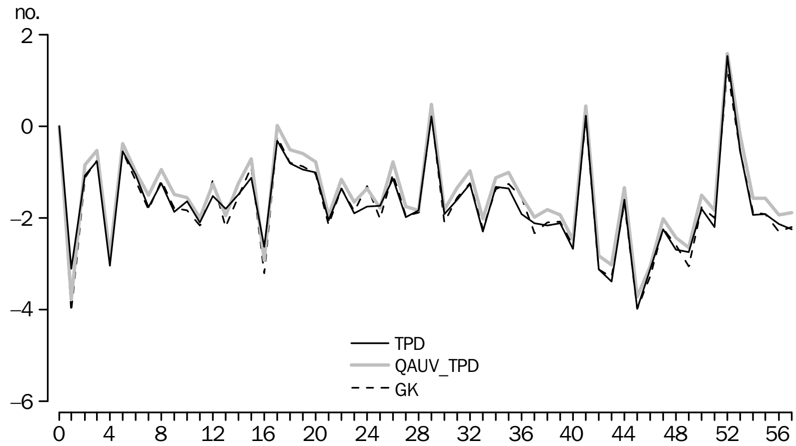


FIGURE 5.3: VEGETABLES EC (FULL WINDOW)

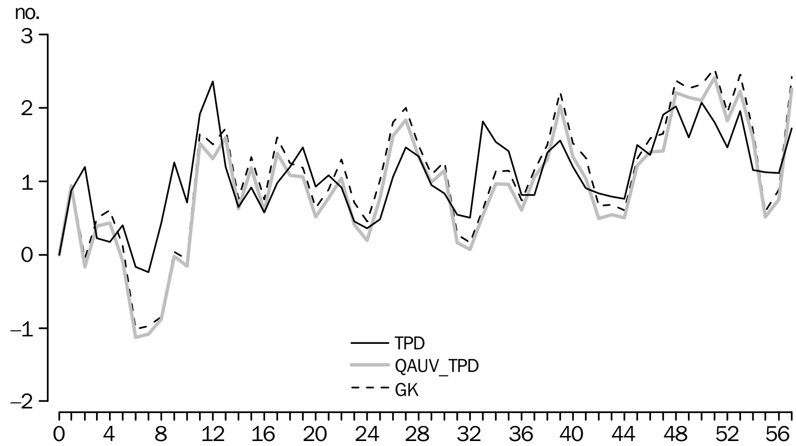
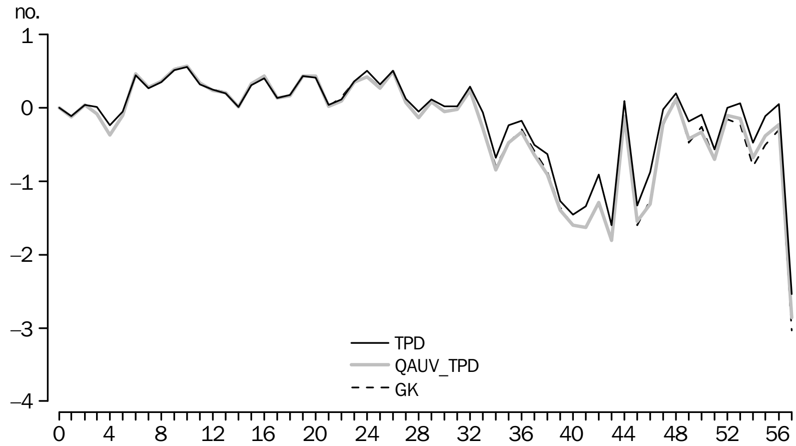


FIGURE 5.4: TOBACCO EC (FULL WINDOW)



EMPIRICAL RESULTS AT THE DISAGGREGATE LEVEL *continued*

FIGURE 5.5: BEEF AND VEAL EC (FULL WINDOW)

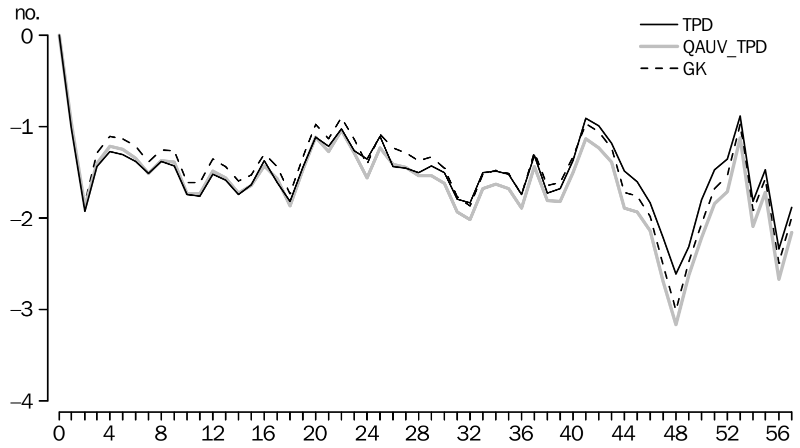


FIGURE 5.6: OTHER NON-DURABLE HOUSEHOLD PRODUCTS EC (FULL WINDOW)

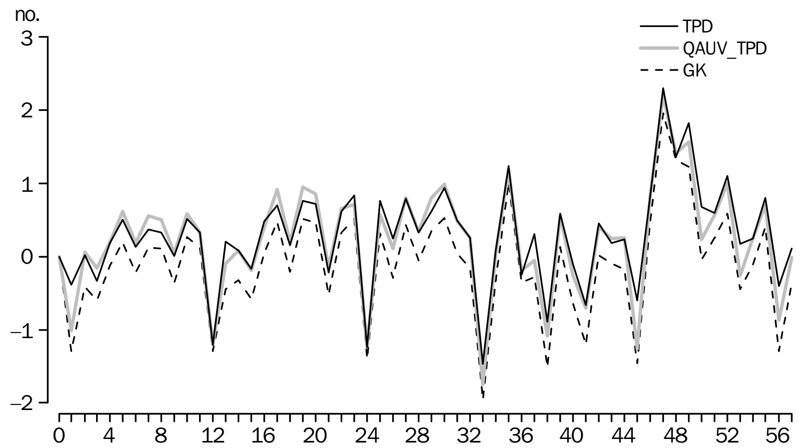
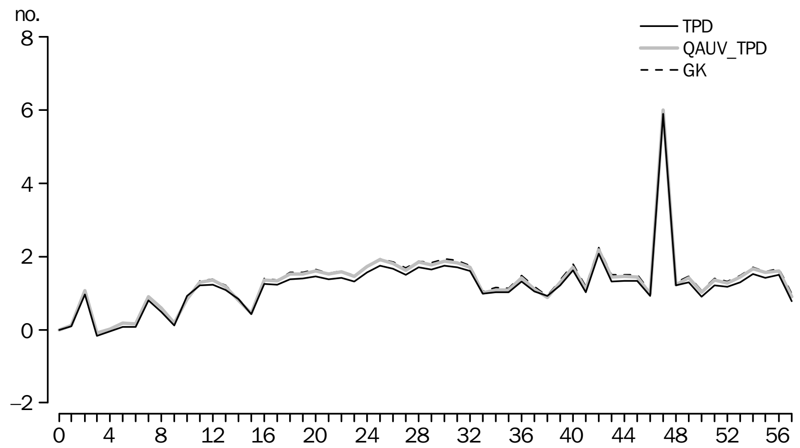


FIGURE 5.7: EGGS EC (FULL WINDOW)



5.10 While the disaggregate results showed small differences among the multilateral methods, certain commodities displayed periods with large differences. An example is presented above in Figure 5.7 for the Eggs EC where the TPD, QAUV_TPD and GK were typically within 2 index points of the GEKS, with the exception of period 47 where the results departed significantly. Using available tools that express GEKS and TPD movements as the product of contributions from individual items, the price fall of the

EMPIRICAL RESULTS AT THE DISAGGREGATE LEVEL *continued*

GEKS (relative to the TPD method) in period 47 is primarily due to a small subset of products experiencing simultaneous falls in price and expenditure. In the following period, these items leave the market at historically low prices and the GEKS index adjusts back to its longer term trend. This example shows that while long-term price trends are similar, there can be significant short-term departures that are driven by the way that each multilateral method uses expenditure/quantity shares to weight products. Similar instances of the GEKS departing from other multilateral methods were observed in other ECs including: Snacks and confectionery; Poultry; and Other non-durable household products.

5.11 With respect to observing behaviour when applying different extension methods, price indexes were constructed for each multilateral method whilst varying the extension methods. Price indexes were then expressed as a difference to the index compiled using the full window of data (e.g. TPD_HS = TPD_HS index less TPD_FULL index) in order to assess the impact of using different extension methods in isolation. Since the full window index for each multilateral method is transitive it was chosen as the benchmark to compare against the different extension methods.

5.12 The results show that the choice of extension method can have a significant impact on the time series. On average, across the extension methods assessed for this publication, the HS method tended to report results closest to those of the full extension for each multilateral method. However, this was not consistent across all ECs, with each extension method performing well for specific commodities.

5.13 Figures 5.8-5.11 below provide an example for the Snacks and confectionery EC. An interesting feature is the high variability of the direct extension method compared to the full window or methods that use a rolling window (i.e. HS, MS, WS) approach. In particular, the magnitude of the difference between direct extension method and the full window results is typically largest immediately following the link month each year (e.g. periods 0, 12, 24 etc). The empirical observation of the direct extension diverging immediately after each link month, then converging toward the full extension result over the year, may be an indication that sparse data (at the beginning of each year) leads to differences in the estimated price indexes. The direct extension method used is also characterised by a shorter maximum window (13 months) compared to the rolling window approaches (25 months) in an effort to replicate the extension method described in Chessa (2016) - this is another possible factor contributing to the differences for the direct extension method.

EMPIRICAL RESULTS AT THE DISAGGREGATE LEVEL *continued*

FIGURE 5.8: SNACKS AND CONFECTIONERY EC (TPD)

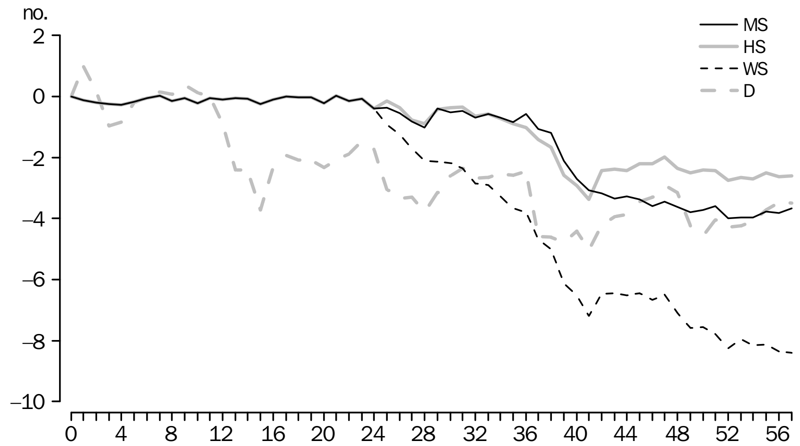


FIGURE 5.9: SNACKS AND CONFECTIONERY EC (QAUV_TPD)

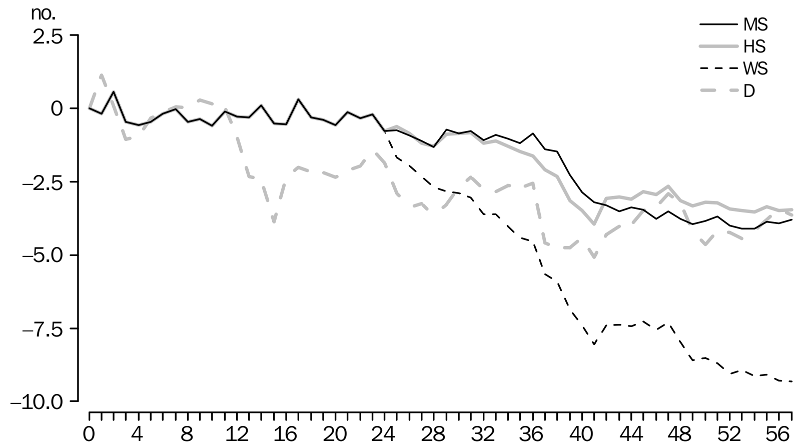
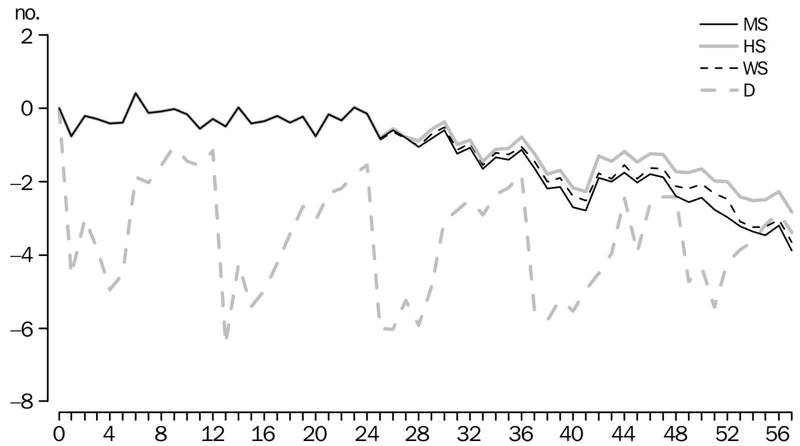
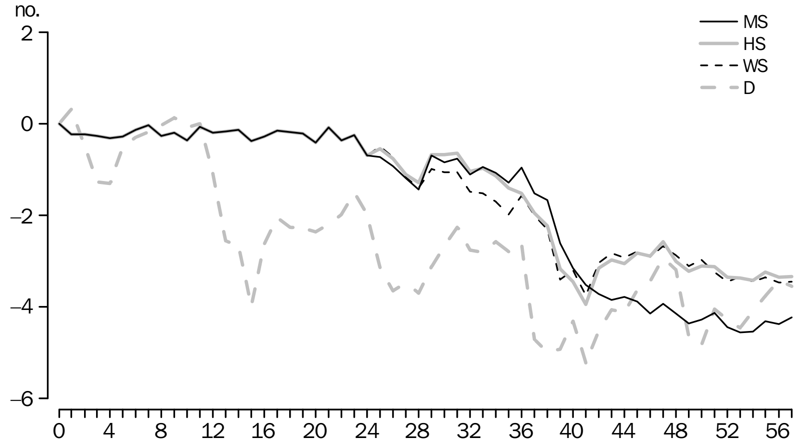


FIGURE 5.10: SNACKS AND CONFECTIONERY EC (GEKS)



EMPIRICAL RESULTS AT THE DISAGGREGATE LEVEL *continued*

FIGURE 5.11: SNACKS AND CONFECTIONERY EC (GK)



5.14 With respect to the rolling window approaches in Figures 5.8-5.11, the indexes are identical for the first 25 periods then begin to diverge as the different splicing methods take effect. It is also apparent that all the extension methods produce a price index that drifts lower relative to results generated when using a full window. For the Snacks and confectionery EC, the HS extension method tends to sit closest to results generated using the full window of data.

5.15 In summary, the empirical evidence at the disaggregate level shows the following features:

i) The multilateral methods assessed in this publication (GEKS, TPD, QAUV_TPD, GK) follow similar price trends over the analysis period. For some commodities, the GEKS departs from the other multilateral methods in the short-term - this is driven by the role of average matched expenditure shares to weight products.

ii) The choice of extension method can significantly impact the time series. The results of this publication lend support for the use of a rolling window approach, specifically the HS, however it was observed that all extension methods tended to drift below the full window benchmark over time. The results for the direct extension method appear to be influenced by the choice of link month.

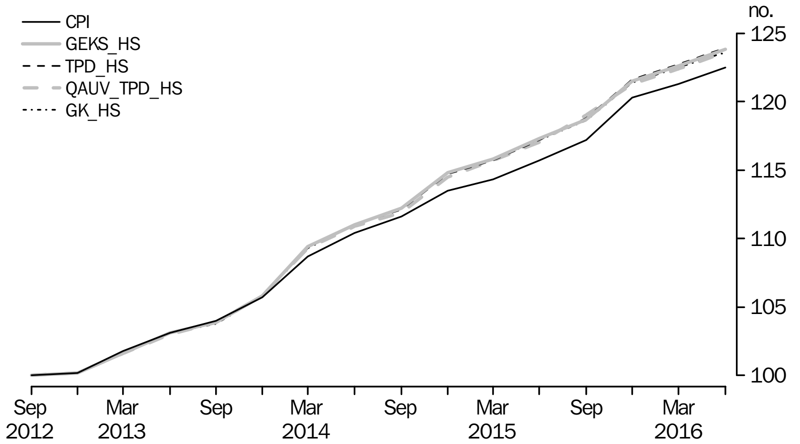
EMPIRICAL RESULTS AT THE AGGREGATE LEVEL

5.16 Empirical results at the aggregate level are presented below for a selection of groups and ECs. The benefit of presenting empirical results that include weighted price movements of both transactions and non-transactions data respondents are to produce an inflation measure that is comparable to the published CPI. The results discussed below are compiled on a quarterly frequency for each individual multilateral method. The HS extension method is used with each multilateral method due to its favourable performance in the above sub-section.

5.17 The empirical results at the aggregate level focus on three main CPI groups where transactions data contribute significantly to the Australian CPI - Food and non-alcoholic beverages, Alcohol and tobacco and Furnishings, household equipment and services (FHES). Figure 5.12 shows the multilateral price indexes rose between 23.6% and 23.9%, while the published CPI rose 22.5% over the time series at the weighted eight capitals with the largest divergences occurring in the Tobacco EC. Comparatively, all the multilateral indexes produced very similar results.

EMPIRICAL RESULTS AT THE AGGREGATE LEVEL *continued*

FIGURE 5.12: ALCOHOL AND TOBACCO GROUP PRICE INDEXES, (weighted eight capitals)



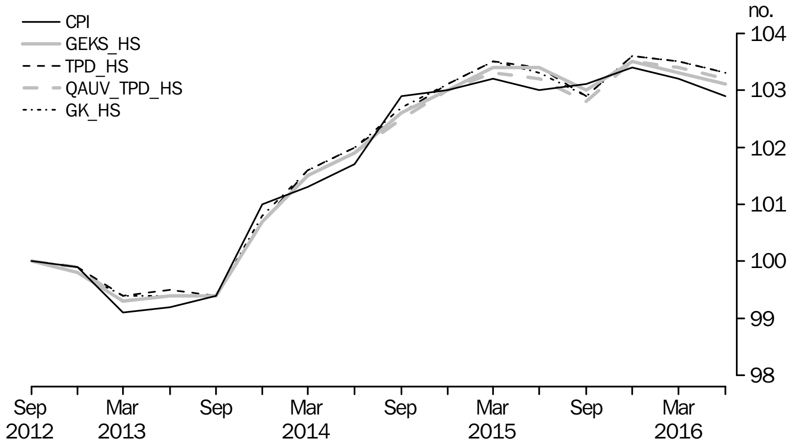
5.18 The divergences between the multilateral price indexes and CPI for the Tobacco EC becomes most noticeable following December 2013, coinciding with the Australian Government significantly increasing the excise charged on tobacco products[8] . The primary reason for the aggregate multilateral indexes to depart from the CPI is the use of expenditure weights at the EA and product level, with the multilateral methods capturing a shift in consumer preferences over time to cheaper tobacco products. Since the tobacco excise is charged on a per cigarette basis (per kilogram for loose tobacco), the multilateral methods are registering larger price rises relative to the CPI as the tax excise has a larger price effect on cheaper (per cigarette/gram) products. The example of tobacco is an interesting study where shifts in consumer preferences have actually caused a higher inflation aggregate, as opposed to traditional substitution behaviour to low inflation goods.

5.19 For the Food and non-alcoholic beverage group the multilateral price indexes rose between 3.1% and 3.3%, while the CPI rose 2.9% over the time series (see Figure 5.13). The multilateral price indexes and CPI show similar trends over time, with some short-term differences predominantly coming from volatile items (e.g. fruit). Across the cities, the relationship between the multilateral indexes and the CPI is mixed, with the CPI sitting higher or lower relative to the multilateral results for certain cities.

8 From December 2013, the Australian government introduced annual 12.5% increases to the tobacco excise, as well as changing the biannual indexation to average weekly ordinary time earnings.

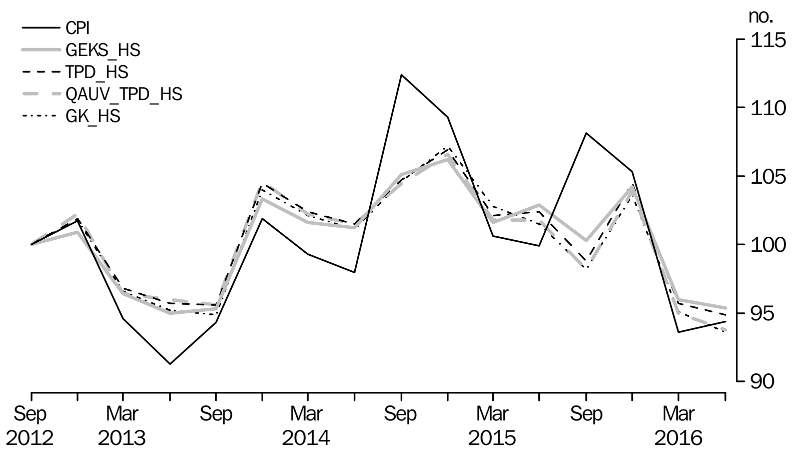
EMPIRICAL RESULTS AT THE AGGREGATE LEVEL *continued*

FIGURE 5.13: FOOD AND NON-ALCOHOLIC BEVERAGES GROUP PRICE INDEXES, (weighted eight capitals)



5.20 Despite reaching a similar price level at the end of the time series (see Figure 5.14), the Fruit EC is the most volatile commodity within the Food and non-alcoholic beverage group, with specific quarters recording significantly different price movements. Of particular interest is the September 2015 quarter, where the CPI rose 8.2% driven by stone fruits and grapes due to seasonal fluctuations in supply. In contrast, the multilateral price index methods registered falls of between 2.5% and 3.6%, with the main contributors consisting of berry products (e.g. strawberries, blueberries). This specific price movement shows the effects of using a fixed quantity index formula at the EA level in the CPI which assumes consumers purchase the same quantity of fruit each period irrespective of relative price change. On the other hand, all multilateral methods (calculated at the EC level) take advantage of the expenditure information available from transactions data and produce price indexes that account for consumers changing their expenditure patterns.

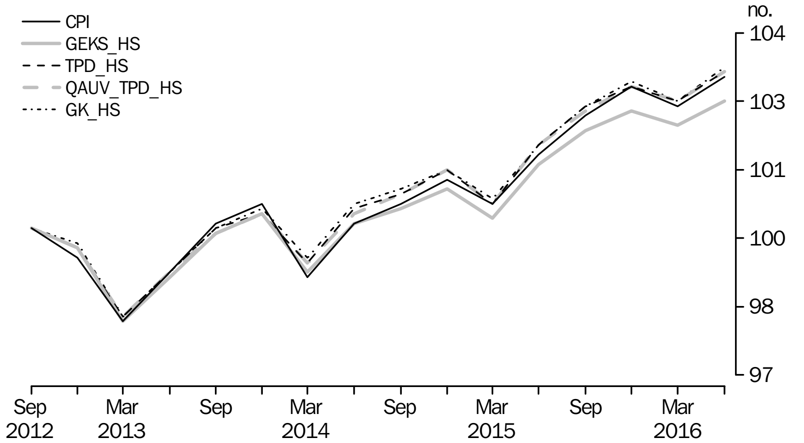
FIGURE 5.14: FRUIT EC PRICE INDEXES, (weighted eight capitals)



5.21 The multilateral price indexes rose between 2.6% and 3.3% for the FHES group, while the CPI rose 3.1% over the time series (see Figure 5.15). Comparing the multilateral methods, the GEKS_HS sits marginally lower than the other methods at the group level.

EMPIRICAL RESULTS AT THE AGGREGATE LEVEL *continued*

FIGURE 5.15: FHES GROUP PRICE INDEXES, (weighted eight capitals)



5.22 In summary, the empirical evidence at the aggregate level shows the following features:

i) The aggregate multilateral methods and published CPI follow similar price trends over the time series. This finding reinforces the traditional practices used by the ABS, since the published CPI is compiled using an unweighted price index formula at the elementary aggregation level.

ii) In instances where the aggregate multilateral methods and published CPI diverge (e.g. tobacco, fruit), it is mainly driven by the multilateral methods using contemporaneous information for weighting purposes that readily capture changes in consumer purchases over time.

CONCLUSION

6.1 The ABS has produced official statistics for Australia for over 100 years to inform decisions on matters of importance. However, the environment in which NSOs operate is changing. Dynamic information is now available, creating new opportunities, and also challenges, for statistical offices. To capitalise on these opportunities, and better meet current and future statistical requirements of the Australian community, the ABS is undertaking a research program to enhance the Australian CPI. This publication focuses on one component of that program, making greater use of transactions data to compile the Australian CPI.

6.2 The Australian CPI currently uses transactions data by calculating an average unit value by product, by taking the quantity and expenditure information over the period of interest and replacing the directly observed price with this unit value. While this has enhanced the Australian CPI, it is recognised that more can be done with big data.

6.3 The availability of timely expenditure data in transactions datasets allows weighted bilateral indexes to be calculated, accounting for consumer substitution. However traditional bilateral methods break down when using transactions data. New methods and processes are required. Temporal multilateral methods have been proposed as they preserve transitivity and make greater use of price and expenditure information.

6.4 Multilateral methods allow NSOs to use the dynamic universe of transactions data to enhance the accuracy of their price indexes. They make greater use of automated processes, providing NSOs with an opportunity to reduce costs across the statistical cycle. Automated processes also provide an opportunity for more timely output - less manual intervention facilitates completion sooner after the reference period, rendering higher frequency price indexes more feasible.

6.5 While having strong theoretical and practical properties, multilateral methods also present NSOs with challenges. Recalculating the index as new periods become available revises the historical series, which is unacceptable for the Australian CPI. To overcome this challenge each new multilateral index must be spliced on to the previously published index levels. Extending the time series in this way impairs transitivity, however this is less severe when using multilateral methods than bilateral methods.

6.6 Another challenge lies in the fact that multilateral methods use prices and expenditure weights for several periods to calculate a transitive index, having potential to affect the relevance of the index. The choice to find a suitable balance for this depends more on the choice of window length and extension method than the multilateral method itself. ABS research has determined that if there is sufficient data available then the estimation window should be a reasonable length, and recommends a 25 month, or 9 quarter window.

6.7 Over the years a range of multilateral methods have been proposed for use in CPI aggregation when using transactions data, while only a handful of NSOs have actually implemented them in their CPI, each using different methods and practices. This stems from a lack of consensus on the best method, as well as the circumstances of each NSO.

6.8 This publication presents a selection of well-known matched-model multilateral methods for producing temporal indexes from transactions data - TPD, GK, QAUV_TPD and GEKS - as well as four different methods for extending the index to create a continuous non-revised time series. With no consensus amongst researchers or leading

CONCLUSION *continued*

NSOs on the best method, the ABS used a framework to help guide the choice of method. This framework was assessed empirically.

6.9 The ABS DQF provides an opportunity to guide the choice of multilateral method by weighting the level of statistical quality at a broad level. Doing so revealed all methods facilitate a better use of resources, as discussed above. From a theoretical perspective, no method satisfied all of the Tests proposed in the literature. The GEKS method has the best established economic properties, but the ABS has not found evidence of substitution bias for the other methods in this context. All methods were found to be adaptable for different data sources, and although they are more complex than bilateral indexes, none is prohibitively difficult to explain.

6.10 Empirical analysis reveals the choice of both multilateral and extension method impacts the estimated price index series. In general, the multilateral methods produced similar trends over time, with short-term departures primarily due to the different use of expenditure shares to weight individual products. Analysis of the extension methods lended support for the HS method, while the direct extension method appeared to be impacted by the choice of link month for certain commodities. Findings also reinforce traditional ABS practices, as aggregate results for the multilateral methods follow similar trends to the published CPI. Where there are divergences they are driven by multilateral methods using contemporary data for weighting purposes, more accurately reflecting changing consumer preferences.

6.11 This publication demonstrates support for the use of multilateral methods as the most opportunistic way to make greater use of transactions data to enhance the CPI. An assessment of the theoretical and practical aspects of the four multilateral methods presented in this publication reveals they are all suitable for use in a CPI context. Whilst having many benefits, multilateral methods present NSOs with challenges which must be considered in a local context. The ABS will continue to assess aspects of multilateral methods as outlined in this publication. The ABS will also consult with peers and experts in order to develop a best practice approach.

6.12 User and stakeholder input is sought to contribute to the resolution of outstanding methodological challenges. A final ABS publication will be released on this topic in mid-2017, which will articulate the precise methods to be implemented in the CPI. These methods will be finalised following extensive consultation with international experts, academics and users.

6.13 For further information relating to making greater use of transactions data in the CPI, readers should write to:

Mr Andrew Tomadini
Director
Consumer Price Index Section
Australian Bureau of Statistics
PO Box 10
Belconnen ACT 2617
E: prices.statistics@abs.gov.au

Country	*Transactions data items*	*Elementary aggregation formula*
Belgium	Supermarket items	Geometric mean
Denmark	Supermarket items	Geometric mean
Iceland	Supermarket items	Geometric mean
Netherlands	Supermarket items	Geometric mean
	Mobile phone and department store items	GK method with a direct annual extension
New Zealand	Audio visual and household appliance items	Imputation Törnqvist Rolling Year GEKS (ITRYGEKS)
Norway	Food medical, retail, petrol and pharmacy items	Geometric mean
Sweden	Supermarket items	Geometric mean
Switzerland	Supermarket items	Geometric mean

ADAPTATION OF MULTILATERAL TESTS FROM VOLUME SHARES TO PRICE AND QUANTITY COMPARISONS

The tests for multilateral indexes presented by Diewert (1999) and Balk (2001) are expressed in terms of volume shares. These volume shares are equivalent to normalised quantity index levels, and can also be expressed in terms of prices, quantities and (multilateral) price levels:

$$Q^t = \frac{\mathrm{p}^{\mathrm{t}}.\mathrm{q}^{\mathrm{t}}/P^t}{\sum_s \mathrm{p}^{\mathrm{s}}.\mathrm{q}^{\mathrm{s}}/P^s} \text{ (A.1)}$$

Where Q^t is the volume share in period t, p^{t} and q^{t} are price and quantity vectors respectively in period t and P^t is the multilateral price level in period t.

Test 1 states that Q^t are positive and continuous and normalised so that

$$\sum_s Q^s = 1 \text{ (A.2)}$$

This implies that the quantity comparison between any pair of entities (e.g. time periods) a and b,

$$Q^{ab} = \frac{Q^b}{Q^a} \text{ (A.3)}$$

is also positive and continuous.

The product test states that

$$V^{ab} = P^{ab} Q^{ab} \text{ (A.4)}$$

$$\Rightarrow P^{ab} = V^{ab} / Q^{ab}$$

where V^{ab} and P^{ab} are the expenditure and price comparisons respectively between periods a and b.

Moreover the expenditure comparison can be expressed as:

$$V^{ab} = \frac{V^b}{V^a} = \frac{\mathrm{p}^{\mathrm{b}}.\mathrm{q}^{\mathrm{b}}}{\mathrm{p}^{\mathrm{a}}.\mathrm{q}^{\mathrm{a}}} \text{ (A.5)}$$

These expenditure comparisons are also positive and continuous. In consequence, the price comparisons are also positive and continuous. For multilateral temporal price comparisons, it is customary to normalise the price index levels such that:

$$P^0 = 1$$

Several Tests, including 2, 2x, 2p, 3, 4, 11, and the first part of Test 8, can easily be expressed in terms of quantity comparisons, and then A.4 and A.5 can be used to derive the implied Test for price comparisons after some simple algebra. In most cases the price comparison Test corresponds closely to the quantity comparison Test, with the possible exception of Test 11. This Test states that combining volume shares additively in each group yields a Fisher quantity index, which means that price levels need to be combined harmonically to yield a Fisher price index.

Derivations of all these tests are not shown here, but Test 3 is presented as an example. Define $\mathrm{P} = \{\mathrm{p}^0, \ldots, \mathrm{p}^{\mathrm{T}}\}$ as the set of original price vectors, $\mathrm{Q} = \{\mathrm{q}^0, \ldots, \mathrm{q}^{\mathrm{T}}\}$ as the set of original quantity vectors, and $\tilde{\mathrm{Q}} = \{\mathrm{q}^0, \ldots, \lambda\mathrm{q}^{\mathrm{c}}, \ldots, \mathrm{q}^{\mathrm{T}}\}$ as the set of quantity vectors after rescaling by a positive constant λ in an arbitrary period c. Then homogeneity in quantities implies that:

ADAPTATION OF MULTILATERAL TESTS FROM VOLUME SHARES TO PRICE AND QUANTITY COMPARISONS *continued*

$$\frac{Q^b(\mathrm{P};\tilde{\mathrm{Q}})}{Q^a(\mathrm{P};\tilde{\mathrm{Q}})} = \begin{cases} \dfrac{Q^b(\mathrm{P};\mathrm{Q})}{Q^a(\mathrm{P};\mathrm{Q})} & (c \neq a,b) \\ \lambda^{-1}\dfrac{Q^b(\mathrm{P};\mathrm{Q})}{Q^a(\mathrm{P};\mathrm{Q})} & (c = a) \\ \lambda\dfrac{Q^b(\mathrm{P};\mathrm{Q})}{Q^a(\mathrm{P};\mathrm{Q})} & (c = b) \end{cases} \quad \text{(A.6)}$$

Using A.5, it is easy to show that:

$$\frac{V^b(\mathrm{P};\tilde{\mathrm{Q}})}{V^a(\mathrm{P};\tilde{\mathrm{Q}})} = \begin{cases} \dfrac{V^b(\mathrm{P};\mathrm{Q})}{V^a(\mathrm{P};\mathrm{Q})} & (c \neq a,b) \\ \lambda^{-1}\dfrac{V^b(\mathrm{P};\mathrm{Q})}{V^a(\mathrm{P};\mathrm{Q})} & (c = a) \\ \lambda\dfrac{V^b(\mathrm{P};\mathrm{Q})}{V^a(\mathrm{P};\mathrm{Q})} & (c = b) \end{cases} \quad \text{(A.7)}$$

Substituting A.6 and A.7 into A.4, it follows that:

$$\frac{P^b(\mathrm{P};\tilde{\mathrm{Q}})}{P^a(\mathrm{P};\tilde{\mathrm{Q}})} = \frac{P^b(\mathrm{P};\mathrm{Q})}{P^a(\mathrm{P};\mathrm{Q})} \quad \text{(A.8)}$$

In other words, Test 3 implies that price comparisons should be unaffected by scaling the quantity vector in any period if prices do not change.

Tests 5, 6, 7 and the second part of Test 8 state that the volume shares are invariant to changing the units of measurement, reordering the time periods, reordering commodities, or aggregating/disaggregating other entities. None of these changes alters expenditure comparisons, so tests that they do not alter volume shares are equivalent to tests that they do not alter price comparisons (this follows from the product Test A.4).

Test 9 states that a time period's influence on quantity comparisons between other periods vanishes as the magnitude of its quantity vector approaches zero. Under the same condition, the time period's influence on price comparisons between other periods must vanish as well. This suggests that Test 9 is inconsistent with Test 3.

Test 10 states that each period's volume share is an increasing function of its quantities, which implies that Q^{ab} is an increasing function of the elements of q^{b}. A corresponding and appealing Test for price comparisons would be that P^{ab} is an increasing function of the elements of p^{b}, but this doesn't seem to follow from Test 10.

Finally, Balk (2001: section 4) clearly describes formulae for the price and quantity indexes under an additive multilateral method, so there is no need to derive these here.

JUSTIFICATION OF NEW TEST RESULTS

Most of the Test results in Table 4.2 are assessed by Diewert (1999) and/or Balk (2001). However, a few combinations of tests and methods are not assessed, including:

- The TPD method's satisfaction of Tests 10, 11 and 12;
- The QAUV_TPD method's satisfaction of any of these Tests;
- Whether these Tests are preserved for extended multilateral indexes.

A brief justification of the corresponding claims in Table 4.2 is appropriate here. We do not present formal mathematical proofs but focus on the reasoning behind each claim; verification of the details is left to the reader.

TPD

The TPD's performance against Tests 10 and 11 is not evaluated here. Test 11 is not critical for this publication, though it seems unlikely that the TPD would satisfy it, given the bilateral TPD index's similarity to the Törnqvist index as opposed to the Fisher index. However, as noted by both Balk (2001) and Rao (2005), the TPD is equivalent to the Rao method, which is not additive (fails Test 12).

QAUV_TPD

The QAUV_TPD satisfies Test 1, since quantity comparisons can be expressed as:

QAUV_TPD continued

$$Q^{ab} = \frac{\pi.\mathrm{q}^{\mathrm{b}}}{\pi.\mathrm{q}^{\mathrm{a}}} \text{ (A.9)}$$

where $\pi = \{v_{i/b}\}_{i=1}^{N}$

and both the quantity vectors and $v_{i/b}$ (derived from a TPD model) are positive and continuous.

If $\mathrm{q}^{\mathrm{t}} = \beta^t \mathrm{q}$ for every t, then from A.9

$$Q^{ab} = \frac{\beta^b}{\beta^a} \text{ (A.10)}$$

so the QAUV_TPD satisfies Test 2x. Similarly, if prices are proportional then the vector π derived from the TPD model satisfies $\mathrm{p}^{\mathrm{t}} = \alpha^t \pi$ for every t, since the TPD satisfies Test 2p (this equality can be verified through examining the equations defining the Rao method, which is equivalent to the TPD). It follows that:

$$P^{ab} = \frac{V^{ab}}{Q^{ab}} = \frac{\dfrac{\alpha^b \pi.\mathrm{q}^{\mathrm{b}}}{\alpha^a \pi.\mathrm{q}^{\mathrm{a}}}}{\dfrac{\pi.\mathrm{q}^{\mathrm{b}}}{\pi.\mathrm{q}^{\mathrm{a}}}} = \frac{\alpha^b}{\alpha^a} \text{ (A.11)}$$

so the QAUV_TPD satisfies Test 2p. As Balk (2001) notes, satisfaction of either Test 2x or 2p is sufficient to guarantee satisfaction of Test 2.

Scaling the quantities in some period does not alter π, as the TPD model uses expenditure share weights. Examination of A.9 reveals that the quantity comparisons satisfy equation A.6 provided π is not altered, so the QAUV_TPD satisfies Test 3.

Similarly, scaling the prices in some period does not alter π, so the QAUV_TPD satisfies Test 4. To verify this, define $\tilde{\mathrm{P}} = \{\mathrm{p}^0, \ldots, \lambda \mathrm{p}^{\mathrm{c}}, \ldots, \mathrm{p}^{\mathrm{T}}\}$; it is easy to show that:

$$s_i^t(\tilde{\mathrm{P}};\mathrm{Q}) = s_i^t(\mathrm{P};\mathrm{Q}) \text{ (A.12)}$$

And using the expression for the reference prices/quality adjustment factors in the Rao method (equivalent to the TPD):

$$v_{i/b}(\tilde{P};Q) = \left[\prod_{t \neq c} \left(\frac{p_i^t}{P_{TPD}^t(\tilde{\mathrm{P}};\mathrm{Q})}\right)^{\frac{s_i^t}{\sum_z s_i^z}}\right] \times \left(\frac{\lambda p_i^c}{P_{TPD}^c(\tilde{\mathrm{P}};\mathrm{Q})}\right)^{\frac{s_i^c}{\sum_z s_i^z}}$$

$$= \left[\prod_{t \neq c} \left(\frac{p_i^t}{P_{TPD}^t(\mathrm{P};\mathrm{Q})}\right)^{\frac{s_i^t}{\sum_z s_i^z}}\right] \times \left(\frac{\lambda p_i^c}{\lambda P_{TPD}^c(\mathrm{P};\mathrm{Q})}\right)^{\frac{s_i^c}{\sum_z s_i^z}}$$

$$= \prod_t \left(\frac{p_i^t}{P^t(\mathrm{P};\mathrm{Q})}\right)^{\frac{s_i^t}{\sum_z s_i^z}}$$

$$= v_{i/b}(\mathrm{P};\mathrm{Q})$$

where the first equality holds due to A.12, and the second equality holds[9] because the TPD satisfies Test 4. The QAUV_TPD satisfies Test 5 because

9 Up to the choice of a TPD normalisation factor, which would cancel out of the QAUV_TPD comparisons.

QAUV_TPD continued

$$Q^{ab}(\hat{\delta}\mathrm{P};\hat{\delta}^{-1}\mathrm{Q}) = \frac{\hat{\delta}\pi.\hat{\delta}^{-1}\mathrm{q}^{\mathrm{b}}}{\hat{\delta}\pi.\hat{\delta}^{-1}\mathrm{q}^{\mathrm{a}}} = \frac{\pi.\mathrm{q}^{\mathrm{b}}}{\pi.\mathrm{q}^{\mathrm{a}}} = Q^{ab}(\mathrm{P};\mathrm{Q}) \quad \text{(A.13)}$$

where $\hat{\delta}$ is an NxN diagonal matrix expressing the change in units of measurement for each commodity. It is easy to show that multiplying each price vector by $\hat{\delta}$ also inflates π in the same way (when expenditures do not change) using logic similar to the justification of Test 4.

The QAUV_TPD method satisfies Tests 6 and 7 because the TPD satisfies these Tests. Similarly, the QAUV_TPD fails Test 9 because the TPD fails this Test, so the prices in a time period continue to influence π even as the quantities sold in that period approach zero.

The QAUV_TPD method's failure of Test 8 derives from the expenditure share weighting used in the TPD: splitting or aggregating across periods in a group alters the weights given to those periods' prices in deriving the π, hence it alters the QAUV_TPD comparisons between other periods.

As with the TPD, the QAUV_TPD is not formally assessed against Tests 10 and 11 here (though it seems unlikely to satisfy Test 11). Its additivity is clear from A.9.

Test performance of extended multilateral indexes

When the methods described in this publication are used to extend multilateral indexes, most price and quantity comparisons are not derived from a single multilateral window but by combining (multiplying or dividing) price movements from a series of partially overlapping multilateral windows. It is important to consider whether Tests satisfied by each multilateral index are preserved after we combine price and quantity comparisons in this way.

Test 1 is clearly preserved after extension - the result of multiplying or dividing positive and continuous functions is necessarily positive and continuous.

Tests 2, 2x and 2p state that if strong conditions hold within a multilateral window, there is a unique price and/or quantity comparison between each pair of time periods. It is easy to see that if these conditions hold in general, they also hold within each multilateral window (or vice versa), and that the unique price and/or quantity comparisons are transitive, so these Tests are preserved after extension.

Tests 3, 4 and 9 relate to the effect of scaling the quantities or prices in one period on the system of comparisons. Clearly, scaling the quantities or prices in that one period cannot affect the system of comparisons in a window that does not contain it. It follows that if these Tests hold within windows containing the period, they hold in general.

Test 5 states that changing any product's units of measurement (across all time periods) should not alter the system of comparisons. So long as any changes in a product's units of measurement are constant across time, this invariance should be preserved after extension.

Test 6 states that reordering the time periods should not alter the system of comparisons, which may be true within each multilateral window but not for an extended index. For instance, suppose we have two windows that overlap in at least two periods (a and b) and that a third period (c) is in one window but not the other. Due to the influence of price and quantity information from period c , the two multilateral windows will in general yield different comparisons between periods a and b . How each extension method forms the multilateral windows and derives the definitive price movement between a and b (from these and/or other windows) depends on the temporal ordering; so reordering the time periods (or changing the way the index is extended) may alter the extended index values.

Test performance of extended multilateral indexes continued

Test 7 states that reordering the commodities should not alter the system of comparisons. Although the methods used to extend the index depend on the temporal ordering, they do not depend on the order of commodities, so this property is preserved after extension.

Additivity means that quantity comparisons can be expressed in the form of A.9. As noted above, most of the quantity comparisons in the extended index are derived by combining (multiplying or dividing) quantity comparisons from individual multilateral windows. It can be shown that, even if the multilateral comparisons within each window are additive, the full system of extended comparisons (products or quotients of additive comparisons) cannot be expressed in the form of A.9 unless the vectors π from every multilateral window (or the quantity vectors in every time period) are proportional to one another. Neither of these conditions holds in general, so Test 12 fails. However, if there is substantial overlap between two windows, the vectors π from each window are likely to be approximately equal, so short term index movements should approximate an additive index closely. Tests 8, 10 and 11 need reinterpretation for temporal comparisons, so the performance of extended indexes against these Tests in their current form is not critical for this publication. That said, it would appear that Test 8 at least does not hold for extended indexes, as aggregating or disaggregating time periods would affect the temporal sequencing that is used in updating the index level; and Test 11 seems likelier to be preserved, based on logic similar to the justification of Test 2 (since the conditions imposed on the price and quantity vectors are similar).

BIBLIOGRAPHY

REFERENCES

Aizcorbe, A., Corrado, C. & Doms, M. 2003. When Do Matched-Model and Hedonic Techniques Yield Similar Price Measures. Working Paper no. 2003-14, Federal Reserve Bank of San Francisco.

Australian Bureau of Statistics (ABS) 2009. ABS Data Quality Framework. cat. no. 1520.0. ABS, Canberra.

Australian Bureau of Statistics (ABS) 2011. Consumer Price Index: Concepts, Sources and Methods. cat. no. 6461.0. ABS, Canberra.

Australian Bureau of Statistics (ABS) 2013. Feature Article: The Use of Transactions Data to Compile the Australian Consumer Price Index. cat. no. 6401.0. ABS, Canberra.

Australian Bureau of Statistics (ABS) 2015. Enhancing the Australian CPI: A roadmap. cat. no. 6401.0.60.001. ABS, Canberra.

Balk, B. M. 2001. Aggregation Methods in International Comparisons: What Have We Learned? ERIM Report, Erasmus Research Institute of Management, Erasmus University Rotterdam.

Chessa, A. G. 2016. A new methodology for processing scanner data in the Dutch CPI. Eurostat review of National Accounts and Macroeconomic Indicators, 1, 49-69.

Collier, I. L. 1999. Comment. International and Interarea Comparisons of Income, Output, and Prices, A. Heston and R. E. Lipsey (eds.), Studies in Income and Wealth, Vol. 61, Chicago: University of Chicago Press.

Diewert, E. W. 1995. Axiomatic and Economic Approaches to Elementary Price Indexes. Discussion Paper No. 95-01; Department of Economics, University of British Columbia.

Diewert, E. W. 1999. Axiomatic and Economic Approaches to International Comparisons. International and Interarea Comparisons of Income, Output, and Prices, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth, Vol. 61, Chicago: University of Chicago Press., 13-87.

van der Grient, H. 2010. How to Analyse Changes in Price Indexes when Applying the RYGEKS Formula. Statistics Netherlands, The Hague.

van der Grient, H. & de Haan, J. 2011. Scanner Data Price Indexes: The "Dutch" Method versus Rolling Year GEKS. Paper presented at the twelfth meeting of the Ottawa Group, 4-6 May 2011, Wellington, New Zealand.

de Haan, J. 2015. A Framework for Large Scale Use of Scanner Data in the Dutch CPI. Paper presented at the fourteenth Ottawa Group meeting, 20-22 May 2015, Tokyo, Japan.

de Haan, J. & Hendriks, R. 2013. Online Data, Fixed Effects and the Construction of High-Frequency Price Indexes. Paper presented at the Economic Measurement Group Workshop, 28-29 November 2013, Sydney, Australia.

de Haan, J. & Krsinich, F. 2014. Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ. Paper presented at the Society for Economic Measurement Conference, 18-20 August 2014, Chicago, USA.

de Haan, J., Willenborg, L. & Chessa, A.G. 2016. An Overview of Price Index Methods for Scanner Data. Paper presented at the thirteenth meeting of the Group of Experts on Consumer Price Indices, 2-4 May 2016, Geneva, Switzerland.

Hill, R.J. 2000. Measuring substitution bias in international comparisons based on additive purchasing power parity methods. European Economic Review, 44, 145-162.

International Labour Organization (ILO) 2004. Consumer Price Index Manual: Theory and practice, International Labour Office, Geneva

REFERENCES *continued*

Ivancic, L., Fox, K. J. & Diewert, E. W. 2009. Scanner data, time aggregation and the construction of price indexes. Paper presented at the eleventh meeting of the Ottawa Group, 27-29 May 2009, Neuchatel, Switzerland.

Ivancic, L., Fox, K. J. & Diewert, E. W. 2011. Scanner data, time aggregation and the construction of price indexes. Journal of Econometrics, 161, 24-35.

Krsinich, F. 2016. The FEWS Index: Fixed Effects with a Window Splice. Journal of Official Statistics, 32, 375-404.

Mackin, C., Oehm, D. & Gow, D. 2012. Quality Management of the CPI collection. Internal report by Statistical Services Branch for the ABS Audit Committee. Canberra, Australia.

Rao, D. S. P. 2005. On the Equivalence of Weighted Country-Product Dummy (CPD) Method and the Rao-System for Multilateral Price Comparisons. Review of Income and Wealth, 51, 571-580.

United Nations Economic Commission for Europe (UNECE) 2016. Report. Proceedings of the thirteenth session of the Group of Experts on Consumer Price Indices, 2-4 May 2016, Geneva, Switzerland.

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** the ABS website is the best place for data from our publications and information about the ABS.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au